REPEATED DNA SEQUENCES FROM THE BOVINE Y CHROMOSOME

by

MARGARET ELLEN MATTHEWS

A thesis submitted in February, 1990 for the degree of Doctor of Philosophy of the Australian National University

STATEMENT

In accordance with the rules of the Australian National University, I wish to state that, except where due reference is made on p. xv and acknowledged within the text, all the work described in this thesis was carried out by myself, under the supervision of Dr. K.C. Reed.

Margaret E. Matthews

ACKNOWLEDGEMENTS

I would like to thank my supervisor Dr Ken Reed for the invaluable experience I have gained while working in this laboratory and for his enthusiasm and support during the pursuit of this project. Dr Klaus Matthaei, David Mann and Dr Mark Bradley have given me much appreciated advice during both the experimental and writing phases of this project which has been much appreciated. The results and implications of my work have been made broader and greatly enriched by the work of Eric Lord on the sheep Y chromosome. The technical assistance and cheerful attitude of Patrick Donoghoe lightened my load considerably as did the helpful attitude shown by Jenny Houghton (particularly in the tedious typing of references) and Alison Cartledge. The support, assistance and friendship given to me by Sandra Beaton contributed a great deal to the successful completion of this thesis.

I would also like to thank the members of the Biochemistry Department who have encouraged me during this period, in particular Dr Maurie Weidemann, Dr Peter Stewart (who also read some of the draft thesis and gave timely advice), Dr Peter Matthews, Dr Fyfe Bygrave and Dr Tony Howells. The companionship and reassurance provided by Peter Kerr and Dr Ann Nestorowicz has also been valued. The steady running of the Department by Bill Nicholson and the other staff under sometimes difficult circumstances has allowed the progress of this research to proceed unhindered by mechanical breakdown. The night shift has been brightened by the presence and occasional chats with Bruce Hickson, who also makes the Department feel more secure in the early hours of the morning. Finally thankyou to Bill Warren, who helped me in an hour of crisis with my worst enemy (the Gene-Master).

The staff in the Hancock library have been very helpful throughout the writing of this thesis, as have the staff at ANU Photographic Services. I would like to give a special thanks to Marie Colville for the care and effort she put into printing the photographs for this thesis.

The ANU has made a considerable effort to facilitate the creation of equal opportunity for women to pursue higher education. I would like to thank the Administration, particularly Colin Plowman, for supporting childcare facilities on campus and all of the members of the Campus Childcare Co-operative for their contribution towards making motherhood and the concurrent pursuit of a PhD a feasible prospect.

Most of all I would like to thank my extended family and my much-loved friends for their tolerance and understanding, especially Eva Bugledich for keeping me sane. None of this work would have been possible without the continuing help, support, guidance and encouragement of my partner Paul Lahiff and the love and patience of our children.

ABSTRACT

Repeated DNA sequences specific to the Y chromosome have been isolated for several species. This observation and the information available from the work of other researchers about the role of the mammalian Y chromosome in sex determination suggested that the presence of sequences which are repeated only on the Y chromosome may be a general rule. Sequences specifically repeated on the bovine Y chromosome would be useful for the diagnosis of sex in cattle embryos, particularly in conjunction with embryo transfer programs. When this work was begun there had been no previous study of the Y chromosomes of ruminants at the molecular level and it was hoped that the isolation of repeated sequences may eventually provide access to coding sequences on this chromosome.

A short repeated sequence, BRY.1, which has a higher copy number in males than females was isolated from a library enriched for male-specific sequences. Unlike the Y-specific repeats found in man and the mouse, BRY.1 is also repeated specifically on the Y chromosome of sheep and goats. This sequence was used to isolate a recombinant phage (EMBL3A.Y1) from a cattle genomic library. This phage contains regions which give a pattern of hybridization typical of high copy repeats which are dispersed throughout the genome, as well as two regions comprised of sequences which are only repeated on the Y chromosome. These two regions were subcloned, analysed by hybridization studies and sequenced. One of these two subclones was then used to reprobe the recombinant phage library and another 19 phage containing Y-chromosomal sequences were isolated. These inserts were mapped and characterized and it was found that many of them, including the original phage, contain sequences which have homology with cDNA probes synthesized from bull testis poly (A+) RNA. The second subclone from EMBL3A.Y1 was used to probe a foetal testis cDNA library and a sequence was recovered which is repeated specifically on the Y chromosome of cattle only. This cDNA is expressed only in the testis, and at a much higher level in the adult than in calf or foetal testis.

The results of this work have lead to a much greater understanding of the structure of the bovine Y chromosome and comparisons with homologues on the sheep Y chromosome have provided a great deal of

information about the evolution of the ruminant Y chromosome. An assay for the presence of the Y chromosome in pre-implantation cattle embryos has been developed and commercialized and variations of the assay are also being applied for the benefit of the livestock industry. The conclusions of this work are discussed in relation to studies of Y chromosome structure and evolution in other species and in relation to the current ideas of general genome evolution.

TABLE OF CONTENTS

TABLE OF FIGURES AND TABLES

ABBREVIATIONS

The following abbreviations are used in this thesis:

bp	base pair(s)
BLOTTO	10% (w/v) skim milk powder (Unigate Diploma), 0.1% (w/v) Na azide
BSA	bovine serum albumin
DMSO	dimethyl sulfoxide
DTT	dithiothreitol
EDTA	ethylenediaminetetra-acetic acid
h	hour(s)
IAC	isoamyl alcohol:chloroform (1:24)
kb	kilobase pairs
LMT	low melting temperature agarose
min	minute(s)
Mya	Million years ago
nt	nucleotide(s)
NZCYM	1% (w/v) NZ amine (Sigma), 0.5% (w/v) yeast extract, 0.1% casamino acids, 85 mM NaCl, 10 mM magnesium sulphate, pH 7.5
PCR	polymerase chain reaction
PE	0.133 M sodium phosphate, pH 6.9, 1 mM EDTA
PEG	Polyethyleneglycol
PEI-cellulose	polyethyl-imine
pfu	plaque forming units
poly (A+) RNA	mRNA molecules having a 3' tail of polyadenylic acid
psi	pounds/square inch
SDS	sodium dodecyl sulfate
sec	second(s)
SM	100 mM NaCl, 10 mM magnesium sulphate, 50 mM Tris-Cl, pH 7.5, 0.1 mg/ml gelatin
SOB	2% bacto-tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 mM KCl, 20 mM Mg^{2+}
SSC	0.15 M NaCl, 0.015 M Na citrate

ssDNA	single-stranded DNA
SSPE	10 mM Na phosphate (pH 7), 0.18 M NaCl, 1 mM EDTA
TAE	40 mM Tris-Cl, 20 mM sodium acetate, 2 mM EDTA, pH7.8
TBE	89 mM Tris-Cl, 89 mM borate, 2 mM EDTA.Na_2, pH 8.3
TE	10 mM Tris-Cl, pH 7.5, 1 mM EDTA
2YT	16 g Tryptone, 10 g yeast extract, 5 g NaCl (pH 7.4) per litre

PUBLISHED PAPERS AND COLLABORATIVE WORK

Chapter 2 of this thesis has been submitted for publication in Cytogenetics and Cell Genetics. The paper is included here in thesis format,under the same title, but with the abstract incorporated into the abstract of this thesis and the reference list combined into the thesis list.

Chapters 2, 3, 4, 5 and 6 describe work done entirely by myself, with the exception of some of the sequencing in Chapter 2 which was done by Klaus Matthaei, the construction of the phage library used in Chapters 3 and 5, which was done by Frances Stewart and the construction of the cDNA library used in Chapter 6, which was done by Sandra Beaton. Sandra Beaton also prepared the poly (A+) RNA used in Chapters 2, 5 and 6 and the Northern blots used in Chapter 6. The sheep probes used in Chapters 3 and 5 were isolated by Eric Lord using cattle probes I had isolated.

Papers and Posters presented at conferences

Matthews, M.E. and Reed, K.C. (1986) A repeated DNA sequence specific to male ruminants. Australian Society for Reproductive Biology, Brisbane.

Matthews, M.E., Matthaei, K.I., Herr, C. and Reed, K.C. (1987) Sex determination of pre-implantation embryos. Australian Society for Reproductive Biology, Sydney.

Lord, E.A., Matthews, M.E. and Reed, K.C. (1987) Y-specific repeated DNA sequences conserved in ruminants. Genetics Society of Australia, Canberra.

Matthews, M.E. (1987) DNA probe technology and early embryos. The Fertility Society of Australia, Sydney.

Patents applied for:

Reed, K.C., Matthews, M.E. and Jones, M.A.S.J. (1986) Sex determination in ruminants using Y-chromosome specific polynucleotides, patent PCT/US87/00254.

Reed, K.C., Lord, E.A., Matthaei, K.I., Mann, D.A., Beaton, S. Herr, C.M. and Matthews, M.E. (1988) Determination of genetic sex in ruminants using Y-chromosome specific polynucleotides, patent PCT/AU89/00029.

CHAPTER ONE

GENERAL INTRODUCTION

CHAPTER ONE

General Introduction

"... the most important inventions (of evolution) are sex and death"
(F. Jacob, 1982)

1.1 The Y chromosome and primary sex determination

Mammalian sexual dimorphism has long been recognized as due to the chromosome constitution XX/XY, where the presence or absence of the Y chromosome determines sex (McCarrey and Abbott, 1979). Primary sex determination is one of the most interesting and accessible problems in the study of mammalian differentiation. The key step in sex determination is the differentiation of the indifferent embryonic gonads. In the presence of the Y-encoded sex-determining gene(s) they develop into testes; in the absence of this gene(s) ovaries develop (Gordon and Ruddle,1981).

The responsible gene(s) on the Y chromosome acts as a dominant trait, that is individuals with the Y chromosome develop testes regardless of how many Y (or X) chromosomes are present. An XO genotype in mammals produces a phenotypic female (Turners syndrome; Turner,1938; cited by Gordon and Ruddle, 1981), so the presence of the Y chromosome is crucial to the determination of a male sexual phenotype.

It is generally agreed that the Y chromosome must have loci for both testis determination and production of H-Y antigen (a male-specific transplantation antigen). H-Y was discovered by Eichwald and Silmser, in 1955, and twenty years later it was proposed that this antigen was the molecule responsible for testis determination (Wachtel *et al.*, 1975). It is now known that H-Y is not likely to be involved directly with testis determination since a strain of mutant mice known as *Sxr'* has been developed whose males lack the antigen (McLaren *et al.*, 1984). On the Y chromosome may be either the structural genes themselves or regulatory sequences acting on autosomal or X-linked structural genes.

The classical view of mammalian sexual differentiation is that a Y chromosome-specific gene, *Tdy* in the mouse (Eicher, 1982), *TDF* in man,

controls the development of the indifferent gonad into a testis. The Leydig cells then secrete androgens which stimulate the development of the male reproductive tract. Sertoli cells secrete Mullerian inhibitory substance which inhibits the development of the female reproductive tract (Jost *et al.*, 1973). In the absence of a testis, the Mullerian duct develops into the Fallopian tubes, uterus and vagina. All subsequent steps in the whole process of sexual differentiation were thought to be hormonally mediated as a consequence of this initial genetic determination of primary sex.

The study of sexual differentiation in marsupials has recently provided evidence that the Y chromosome directs the development of some secondary sexual characteristics independent of the action of testosterone (O *et al.*, 1988). At birth, Tammar wallabies show no gonadal differentiation and there is no evidence of hormonal secretion by the male gonads but there are already clear signs of sexual dimorphism, such as scrotal bulges in males and mammary and pouch anlagen in females.

There are also examples of apparent sex differences which precede gonadal differentiation in eutherian mammals. Blastocoele formation is faster, and somite number is greater, in male mouse embryos at day 9 of gestation, compared with female embryos (Sellar and Perkin-Cole, 1987). Male rat embryos are heavier and contain more protein on day 12 of gestation (Scott and Holsen, 1977). It has also been observed that even at the 8-cell stage cytotoxic antibodies raised specifically to male cells can kill XY but not XX mouse embryos (Krco and Goldberg, 1976; Shelton and Goldberg, 1984). Such sexual dimorphisms, preceding gonadal differentiation, are independent of gonadal hormone secretion and must be due to the effects of other sex-linked genes. It is possible that sex-linked genes could control the rate of cell division, during early embryonic events, thus conferring a sex-specific growth advantage on the male embryo (Heslop *et al.*, 1989).

1.2 The role of other chromosomes in sex determination

Accumulating evidence (Eicher and Washburn, 1983) indicates that, in addition to the Y chromosome, genes on autosomes are also involved in primary sex determination. For example, when the Y chromosome of *Mus musculus domesticus* was introduced into the genetic background of the C57BL mouse strain, XY offspring developed as true hermaphrodites or females. The formation of ovarian tissue in these animals could be

interpreted as the result of 'miscommunication' between the *M.m. domesticus*-type Y chromosome and C57BL testes-determining genes on other chromosomes. Three such autosomally located genes *Tda-1*, *Tda-2* and *Tas*, have been identified (Eicher and Washburn, 1986). It has been suggested that *Tda-1* responds poorly to the signal from the foreign Y, creating a fertility barrier between two populations which could lead to species' diversification.

In the mouse, chromosome 17 in particular seems to have loci involved in sex determination (Erickson *et al.*, 1987). The proximal region has loci affecting spermatogenesis, including the transmission ratio-distorting properties (Erickson *et al.*, 1981) and male sterility properties (Hammerberg, 1981) of the t-complex. The hybrid sterility locus (*Hst*) is on proximal chromosome 17 but has an effect on spermatogenesis which is not related to t-complex alleles (Forejt, 1981). It is worth noting that the major histocompatibility complex of the mouse is also on chromosome 17.

There is also a suggestion (de la Chapelle, 1987) that in the human the presence or absence of a Y-linked gene(s) is not the only factor involved in testis determination. Most XX males have been shown to carry some material from the Y chromosome assumed to include the Testis Determining Factor (*TDF*), but Y-derived sequences have not yet been found in some XX males and hermaphrodites and there may be familial inheritance of sex-reversal. This led to the postulation of the existence of an autosomal gene *TDFA* (de la Chapelle, 1987), perhaps homologous to one of the autosomal loci described for mice.

1.3 Sex determination in non-mammalian systems

Since the process of sexual differentiation and the steroid sex hormones are the same throughout the vertebrates it would seem that the genetic sex determining mechanisms would have been stabilized at an early stage in evolution but this is not the case. Autosomal loci are usually involved in sex determination in fish, while in some fish and amphibia either sex may have heterogametic sex chromosomes (Ohno, 1967). In birds the female is the heterogametic (WZ) sex. Many snakes and lizards have heteromorphic sex chromosomes but they are rare in turtles and tortoises and absent in crocodiles (Bull, 1980). Some reptiles with distinct sex chromosomes have a Y in the males of some species, a W in the female

of others. In reptiles which lack detectable sex chromosomes, the sex of offspring is determined by environmental factors (Charnov and Bull, 1977).

Deeming and Ferguson (1988) studying sex determination in the alligator, proposed a model in which gonadal growth is influenced by temperature. According to this model, a threshold gonadal size must be reached by a particular time in development to ensure that the gonad can develop as a testis. At low incubation temperatures gonadal growth is slow and the gonad fails to reach a threshold size by the critical stage of development, so it becomes an ovary. At intermediate temperatures the gonad is near the threshold size at the critical time so that some males and some females result.

In reptiles, temperatures which cause the embryo to differentiate and develop fastest are also the ones that cause them to become male. Deeming and Ferguson (1988) postulated that this is the optimal temperature for transcription, translation and enzyme activities and the dose of the sex determining gene product will be maximal at this temperature, leading to male differentiation. One hypothesis (Deeming and Ferguson, 1988) is that a DNA-binding protein encoded by the sex-determining gene may regulate gene expression at both gonadal and extra-gonadal sites, and may in turn control the release of various regulatory hormones from the hypothalamus.

1.4 Sex determination and the rate of growth

To explain the use of such varying methods of sex determination by vertebrates, Mittwoch (1986) has advanced the theory that the likely common factor in sex determination in vertebrates is the regulation of growth. The mammalian Y chromosome must have loci which somehow direct the gonads to develop into testes. A mechanism that depended on temperature would not be feasible in mammals so mammals adapted one of the previously evolved sex chromosome mechanisms which made sex determination independent of the environmental temperature. According to this theory the relevant sequences of the Y act as a growth enhancing factor which has a special effect on the somatic cells of the gonad. The presence of the Y chromosome accelerates the growth of the gonadal rudiment which consequently develops into a testis. The testes and ovaries differentiate from a potentially hermaphroditic rudiment (Mittwoch, 1986). The components for making both gonads are the same but the indifferent

period, during which the gonad appears undifferentiated, is very short in the male. While the process of testicular differentiation is definitely under the control of the Y chromosome, since it tips the balance in favour of a growth rate fast enough for testis development, the relevant DNA sequences do not have to be unique to the Y chromosome. Heslop *et al.* (1989) proposed that the serologically determined male antigen (H-Ys) may have a function in growth regulation and that this antigen may play a role in the differential gonadal growth rates suggested by Mittwoch (1986). McLaren (1988) suggests that in mammals a Y-linked gene acts as a switch to activate sex-determining genes likely to be located on the autosomes or the X chromosome rather than the Y chromosome.

1.5 Y-chromosomal genes

The first recognizable sign of testicular differentiation seen in the rat gonad is the appearance of Sertoli cells which join each other and encompass the germ cells, initiating formation of the seminiferous tubules and leading to the development of the testis. If there is no Y chromosome, development is slow, the Sertoli cells fail to form and the germ cells continue to divide, then initiate processes which eventually lead to the formation of follicles. Early models postulated that the differentiation of the Sertoli cells was due to the action of exogenous factors, either through cell-cell contact or through a diffusible substance (McLaren, 1987). The finding that Sertoli cells in XX/XY chimeric mouse testes are exclusively XY (Singh *et al.*, 1987), has led to the proposal that the Y chromosome acts cell-autonomously to bring about Sertoli cell differentiation. In this model, all other aspects of foetal testicular development are triggered by the Sertoli cells without further Y-chromosomal involvement, and no diffusible testis-determining molecule is produced. A third possibility is that sex commitment involves an interaction between a cell-autonomous and an exogenous factor (McLaren, 1987).

Whatever the male determining factor carried by the Y chromosome, its genetic analysis has been impaired by its unique haploid state. Unlike the other nuclear chromosomes, the Y has little opportunity to recombine with a homologue, making genetic linkage studies of the Y chromosome difficult if not impossible. This at least partially accounts for the relative dearth of genes mapped to the Y chromosome of the mouse (Roderick *et al.*, 1984) or human (McKusick, 1983). Historically, attempts to establish the Y-

linkage of certain traits have been inconclusive because of the difficulty of distinguishing true Y-linked inheritance from sex-linked expression (Morton, 1957). Nonetheless there is strong evidence for a number of genes on the Y chromosome in addition to the male determinant(s). It has been argued (Burgoyne, 1987) that a minimum of four loci are involved in various Y-chromosomal effects.

Male mice with the XX karyotype have been identified (Cattanach *et al.*, 1971). The trait is dominant and the responsible locus is designated *Sxr* (sex-reversal). Singh and Jones (1982) demonstrated that in *Sxr* mice a small segment of the Y chromosome was transferred to the X. XX *Sxr* mice develop testes, therefore the transferred segment should contain a gene(s) responsible for primary sex determination. The testes are small and functional sperm are absent. Eicher and co-workers (1983) described another mutant in which a substantial amount of the Y chromosome is transferred to the X. This mutant develops testes containing functional sperm. It can be speculated that the Y-chromosomal region present in this mutant, but absent from the XX *Sxr* mice contains loci necessary for sperm maturation (*Spy*; Levy and Burgoyne, 1986). An *Sxr* derivative, *Sxr'*, does not express the male antigen as detected by transplantation studies, H-Yt, and has also lost the *Spy* function (Burgoyne *et al.*, 1986), so H-Yt antigen may mediate the *Spy* function, or the loci may be identical.

Bishop and Hatat (1987) have identified a family of testis specific RNA transcripts encoded by the mouse Y chromosome. These cDNA clones were isolated using a Y-specific probe from a flow sorted mouse Y chromosomal library. This work shows that the mouse Y chromosome is transcriptionally active in the male reproductive organ. The transcripts are from a region of the Y outside the critical Sxr region.

Y-chromosomal material has also been reported in human XX males (Guellaen *et al.*, 1984; Page *et al.*, 1984). Most XX males occur because of terminal exchange of sequences between the X and Y chromosomes, resulting in the transfer of the sex determining gene *TDF* to the X chromosome (Petit *et al.*, 1987). The differing amounts of Y-derived sequences in the genome of XX males has enabled the construction of deletion maps of the short arm of the Y chromosome (Affara *et al.*, 1986;

Geldwerth *et al.*, 1985). Such maps place the male sex determining gene(s) in a sub-telomeric position.

In 1987 Page and his co-workers reported the cloning and sequencing of a gene from the sex-determining region of the human Y chromosome which encodes a zinc-finger protein and was later designated *ZFY*. *ZFY* may be the gene for the testis determining factor (*TDF*). The X chromosome carries a homologue, *ZFX*, and these loci will be discussed in more detail in the next section.

Two genes have been assigned to the proximal portion of the long arm of the human Y chromosome (Yq): a gene affecting spermatogenesis (Tiepolo and Zuffordi,1976), and a gene affecting height and tooth size (Alvesalo and de la Chapelle, 1981). There may also be a regulatory influence of the Y chromosome which prevents the manifestation of Turner stigmata in males (Buhler, 1980). The H-Y locus may be on Yq (Simpson *et al.*, 1987), which raises the possibility that it may be identical to the spermatogenesis gene (Craig *et al.*, 1987). Of the several antigens which have been called H-Y (Wiberg, 1987), it seems likely that it is the one detected by female cytotoxic T-lymphocytes in vitro, H-Yc, which is on Yq in man (Simpson *et al.*, 1987), as opposed to the antigen defined by serological techniques (H-Ys). Recently the gene for a male-enhanced antigen (*MEA*), has been cloned. This gene maps to chromosome 6 of man and chromosome 17 in the mouse (Lau, 1987). *MEA* is a candidate for H-Ys (Lau, 1989).

In humans the sex determining gene(s), *TDF*, is assigned to the short arm of the Y chromosome. The rest of the chromosome is mostly constitutive heterochromatin (permanently condensed DNA), believed to consist of highly repeated sequences which are probably genetically inert (Buhler,1980). However there is evidence of a single-copy sequence in Yq (long arm) (Vergnaud *et al.*, 1986), and reports of other single-copy sequences on distal Yq (Wolfe *et al.*, 1984; Cooke *et al.*, 1984). Figure 1.1 is a schematic diagram of the human Y chromosome, showing the approximate location of *TDF* and other regions and loci referred to in the text.

Associations between the Y chromosome and autosomes occur frequently in man and the pygmy chimp (Weber *et al.*, 1988) and are presumed to be due to the mutual attraction between constitutive heterochromatin on the chromosomes involved. The addition of satellite DNA from other chromosomes does not interfere with fertility in either species. Likewise pericentric inversion of the Y chromosome, which is rare (Schmid *et al.*, 1984), does not influence the expression of male-determining factors or the genes required for normal spermatogenesis.

1.6 Y-chromosomal origins and homology with the X

The X and Y chromosomes are thought to have evolved from an identical pair of chromosomes (Ohno, 1967), but the human X is now three times larger than the Y chromosome (Goodfellow *et al.*, 1985). At least thirty genes have been found to be X-linked in mice (Roderick and Davisson, 1981), with no evidence for corresponding loci on the Y chromosome. In man over 200 genes have been assigned to the X (Miller *et al.*, 1984), with little or no evidence of corresponding loci on the Y chromosome. The need to retain the sex-determining gene(s) on the Y chromosome is a strong theoretical reason for the absence of recombination between the sex chromosomes (Goodfellow *et al.*, 1985) and the large number of X-linked genes not on the Y chromosome testifies to the absence of extensive sex chromosome recombination.

One of the consequences of an XX/XY sex determination system is that females have two copies of X-linked genes while males have only one. This difference in dosage has been overcome in mammals by the inactivation of one of the X chromosomes in the somatic cells of normal females (Lyon, 1972). The classic model of sex chromosome evolution from homologous chromosomes places cross-over suppression between the ancestral pair as the primary event, with later accumulation of differences by mutation (Erickson and Goodfellow, 1984). In this model, X-chromosome inactivation occurs later in evolutionary history to maintain the gene dosage of genes now only on the X. There is evidence from a comparison of sex chromosome differentiation and X-inactivation in monotremes, marsupials and placental mammals (Graves, 1987), that a progressive reduction of the Y chromosome has been accompanied by a gradual spreading of inactivation into newly unpaired regions of the X chromosome.

Chandra (1985) has argued that the X chromosome carries a testis-determining gene which must be inactivated in females, leading to a reduction in the effective copy number of sex-determining genes, whose continued activity in XY cells leads to maleness. A similar theory was proposed by German (1988). In this 'dosage hypothesis of gonadal dimorphism' there are homologous gonad-differentiation loci (GDLs) on the X and Y, which do not normally recombine at meiosis. When two Xs are present, one is inactivated so that only one GDL is transcribed, but when the X and Y are present the abundance of the product from two GDLs leads to testis determination. According to this argument, X chromosome inactivation has a function in sex determination and is not merely an adaptation to an XY sex determination system. Both theories fit well with the hypothesis that the X and Y were originally homologous, with the GDLs, or *TDX* (Testis determining factor on the X chromosome) and *TDY*, being alleles. Both theories also remove the need to postulate a mutation on one chromosome that led to testicular rather than ovarian development. These ideas might also explain why the *'TDF'* locus has been so elusive, since workers have been searching for a Y-specific sequence which may also be on the X.

The wood lemming *Myopus* has provided the most conclusive evidence that the X chromosome is involved in testis determination (Fredga, 1988), with sex ratios biased towards the female and two types of fertile female: XX and XY. It seems that in *Myopus* and in the lemming *Dicrostyonx*, some of the XY animals function as fertile females due to an X-linked gene repressing the male-determining effect of the Y.

One of the major limitations on the differentiation of heteromorphic sex chromosomes is that sex chromosome pairing must occur at meiosis for normal chromosome disjunction. The genes in the pairing segment of the X, which is not inactivated, represent some of the still functionally paired, non-dosage compensating, genes of the ancestral autosome pair (Polani, 1982). Pairing between the tip of the X chromosome short arm and Y chromosome short arm at meiosis has been seen and the telomeric association between the X and Y chromosomes implies sequence homology (but see Ashley, 1984).

Extensive homologies have also been found outside the pairing region of the X and Y chromosomes (Page *et al.*, 1984; Cooke *et al.*, 1984; Koenig *et al.*, 1984; Daiger *et al.*, 1982), and there is evidence at least some of this homology is due to relatively recent transposition of material from the X to the Y chromosome (Page *et al.*, 1984; Koenig *et al.*, 1985). Random sequences isolated from Y chromosomal cosmid libraries also seem to undergo meiotic exchange between the X and Y chromosomes (Cooke *et al.*,1985; Simmler *et al.*, 1985). Bishop *et al.* (1984) have shown with human Y chromosome probes that sequences having homology between the X and Y chromosomes represent about 40% of the euchromatic (non-compacted) portion of the human Y. It is possible a substantial part of the human Y chromosome at least, is made up of rather recently acquired sequences. At least one sequence (*DXYS1*), shared by the human X and Y chromosomes (Figure 1.1), is found on the X and not on the Y chromosome in hominoid apes and thus appears to have been transposed from the X to the Y chromosome since the divergence of human from chimpanzee (Page *et al.*, 1984).

Actin sequences are dispersed throughout the genome and are found on both the X and Y chromosomes in man (Heilig *et al.*, 1984). On the X chromosome the sequences are in the pericentric (euchromatic) region of the long arm and the Y-linked sequences have been localized immediately proximal to the heterochromatic region (*ACT2* - shown in Figure 1.1). Both the X and Y linked actin sequences are pseudogenes, that is sequences which share homology with a coding sequence but never actually produce a message that is translated to give a protein product. The pericentric location of the X-linked actin does not correspond to the recently observed region of X-Y homology, and the presence of the pseudogenes may not reflect homology between the sex chromosomes (Heilig *et al.*, 1984). Several human actin pseudogenes have been characterized which probably arose by reverse transcription of mRNAs coding for a cytoskeletal-type actin (Moos and Gallwitz, 1983).

Sequences cross-reacting with a probe for the argininosuccinate synthetase gene are also present on many human chromosomes including the X and Y chromosomes (Daiger *et al.*,1982). These X and Y linked sequences are also pseudogenes (Freytag *et al.*, 1984).

It has been suggested that recombination occurs in the pairing region shown in Figure 1.1 (Burgoyne, 1982). Genes in this region show pseudoautosomal inheritance: only partial sex linkage if occasionally exchanged, no sex linkage if frequently exchanged, and there is a gradient of recombination of Y-specific sequences (Goodfellow *et al.*, 1987). The pseudoautosomal telomeres combine with a frequency of 50% (Weissenbach *et al.*, 1987), therefore the human sex chromosomes cross over at each male meiosis.

The gene for the enzyme steroid sulphatase (*STS*), is the only well-characterized X-linked locus in humans which has no functional, Y-linked equivalent and yet escapes X-inactivation (Craig and Tolley, 1986). There is no evidence for a functional Y-linked *STS* allele in man but the *STS* gene in the mouse behaves pseudoautosomally with an apparently functional homologue on the Y chromosome (Keitges *et al.*, 1985). The simplest explanation for the apparent differences in the organization of the *STS* gene in mice and man is that the human Y has rearranged since their divergence and represents a rearranged form of an ancestral sex chromosome organization in which both X and Y alleles were functional and whose identities were maintained by crossing over (Craig *et al.*, 1987).

In humans the blood cell antigenic determinant 12E7 (Goodfellow *et al.*, 1983) is controlled by the X-linked gene MIC2, which is in the same region as the X-linked *STS* gene. There is an additional Y-linked 12E7 controlling locus (Figure 1.1) in the euchromatic region of the short arm of the Y chromosome (Buckle *et al.*, 1985). This was the first translated product of the Y chromosome to have been identified. In addition these sequences were the first to have been shown to be shared by the sex chromosomes in the pairing region.

The sharing of genes by the sex chromosomes might reflect a common origin or might be due to a recent exchange of genetic material. The pseudoautosomal gene MIC2 only rarely recombines between the sex chromosomes (< 5%) and based on the elevated recombination events in the pseudoautosomal region, Goodfellow *et al.* (1986) predicted this gene would lie close to the Y-specific region encoding the sex-determining gene. A meiotic map of the human pseudoautosomal region was constructed using DNA probes isolated at random from Y-chromosome genomic libraries and

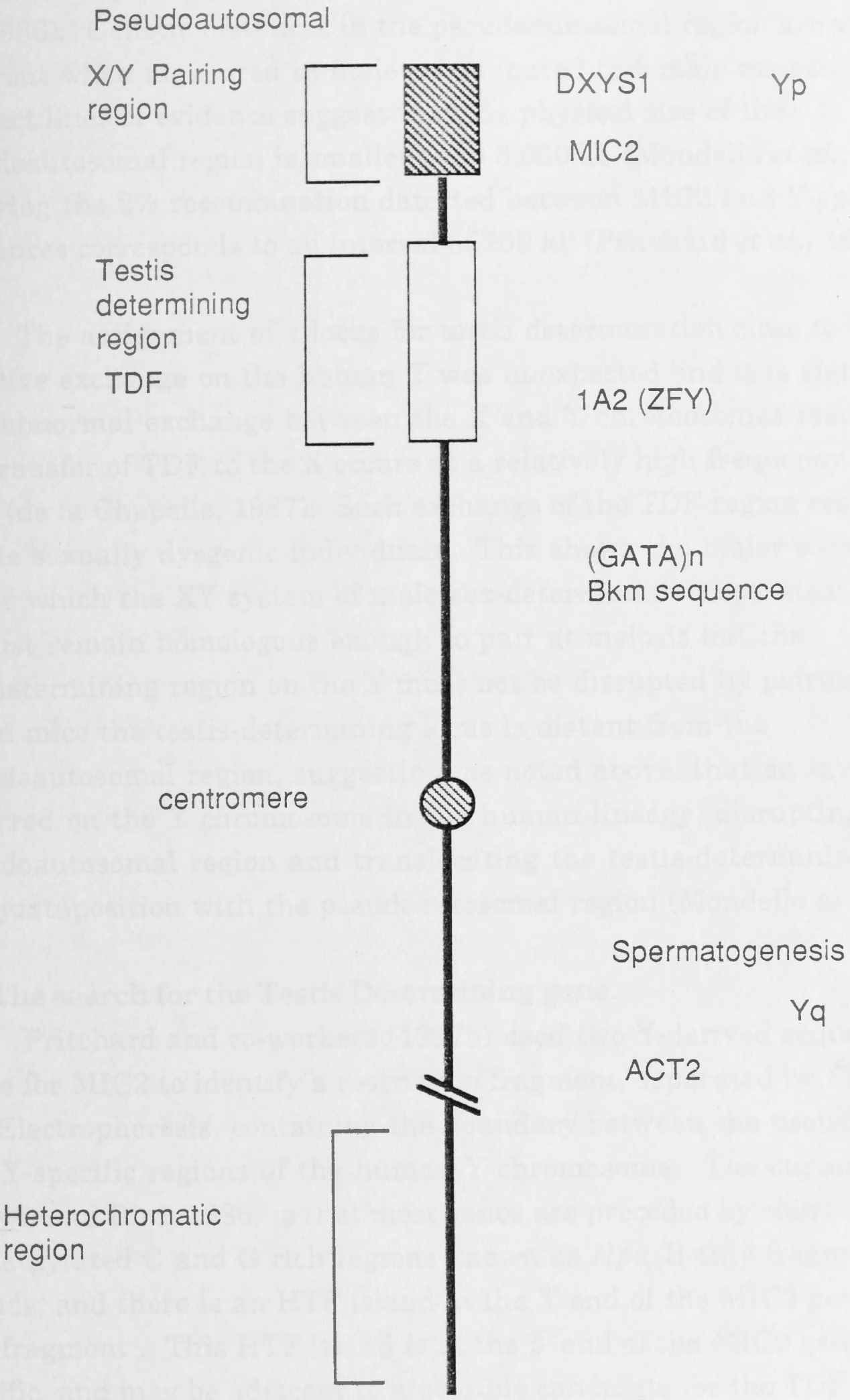


Figure 1.1. Schematic diagram of the human Y chromosome showing regions and loci referred to in text. Adapted from Muller (1987).

probes corresponding to the gene MIC2 (Goodfellow *et al.*,1986; Rouyer *et al.*, 1986). Genetic distances in the pseudoautosomal region are very different when measured in male as compared to female meiosis. Several indirect lines of evidence suggest that the physical size of the pseudoautosomal region is smaller than 5,000 kb (Mondello *et al.*, 1987), implying the 2% recombination detected between MIC2 and Y-specific sequences corresponds to an interval of 200 kb (Pritchard *et al.*, 1987a).

The assignment of a locus for testis determination close to the region of active exchange on the human Y was unexpected and it is significant that abnormal exchange between the X and Y chromosomes resulting in the transfer of TDF to the X occurs at a relatively high frequency (approx. 10^{-4}) (de la Chapelle, 1987). Such exchange of the TDF region results in sterile sexually dysgenic individuals. This shows the major constraint under which the XY system of male sex-determination operates: the X and Y must remain homologous enough to pair at meiosis but the sex-determining region on the Y must not be disrupted by pairing with the X. In mice the testis-determining locus is distant from the pseudoautosomal region, suggesting, as noted above, that an inversion occurred on the Y chromosome in the human lineage, disrupting the pseudoautosomal region and translocating the testis-determining region into juxtaposition with the pseudoautosomal region (Mondello *et al.*, 1987).

1.7 The search for the Testis Determining gene

Pritchard and co-workers (1987b) used two Y-derived sequences and a probe for MIC2 to identify a restriction fragment, separated by Pulsed Field Gel Electrophoresis, containing the boundary between the pseudoautosomal and Y-specific regions of the human Y chromosome. The current opinion (Brown and Bird, 1986) is that most genes are preceded by short unmethylated C and G rich regions known as *Hpa* II tiny fragments (HTF) islands, and there is an HTF island at the 3' end of the MIC2 gene within this fragment . This HTF island is at the 3' end of the MIC2 gene, is Y-specific, and may be adjacent to a possible candidate for the TDF gene on the human Y.

Page *et al.* (1987) cloned a 230 kb region of the human Y chromosome spanning the entire deletion of an XY woman, interval 1A2 (Figure 1.1). An HTF island was found near the distal end of the cloned region, which is

probably the same island as that identified by Pritchard *et al.* (1987b). Some of the sequences within this region are highly conserved, with homologous sequences present in the DNA of all mammals examined. One highly conserved region was sequenced and seems to encode a DNA-binding protein, with 13 'zinc-finger' domains (first described in frog transcription factor IIIA: Miller *et al.*, 1985; Brown *et al.*, 1985), that could act as a transcription-regulating factor, perhaps switching on the male differentiation pathway. Page (1988) later designated this sequence *ZFY*. In most species of placental mammals there are two *ZFY*-related loci but the mouse has four (Page 1988). *Zfy-1* and *Zfy-2* map to the sex-determining region of the mouse Y, *Zfx* is on the mouse X, but the fourth locus is on chromosome 10 (Mardon and Page, 1989; Nagamine *et al.*, 1989). It is likely the human *ZFY* and *ZFX* are true homologues that share a common ancestral gene, but it is unlikely that either is a pseudogene since both show a large degree of conservation among mammals (Page *et al.*, 1987). If *ZFY* is the testis-determining factor, the presence of a homologue on the X chromosome would seem to support dosage compensation theories, such as those of Chandra (1985) and German (1988), and could also be reconciled with ideas of a threshold amount of some substance producing testicular differentiation (McLaren, 1987).

The interval of the Y which contains *ZFY* (1A2) is absent in some human XX males and XX hermaphrodites and is at least grossly intact in many XY females (Page *et al.*, 1987). Sex reversal in such cases may be due to mutations in autosomal or X-linked genes, whose products function together with, or downstream of *TDF* in the sex determination pathway. In mice both *Zfy-1* and *Zfy-2* are present in XX *Sxr* male mice. *Zfy-1* is present but *Zfy-2* is absent in XX *Sxr'* male mice; therefore *Zfy-1* and *Zfy-2* are not both necessary for testis determination (Mardon *et al.*, 1989; Nagamine *et al.*, 1989). Both *Zfy-1* and *Zfy-2* are transcribed only in mouse adult testis, with no *Zfy* transcripts being detected in female tissues or in whole male embryos or in dissected newborn male tissues (Mardon and Page, 1989). This suggests that *Zfy* may have a role in reproduction in the adult male.

ZFY is not on the X or the Y in marsupials (Sinclair *et al.*, 1988). This implies that *ZFY* is not the primary sex determining gene in marsupials, so that either the genetic pathway of sex determination in marsupials and eutherians differs, or *ZFY* is not the primary signal in

eutherians. In the Tammar wallaby Page's clone maps to two autosomes, including one which has the loci *OTC* and *DMD* which are X-linked in eutherians (Sinclair and Graves, 1988). This suggests that in a marsupial ancestor, a large portion of the X was translocated to the autosomes. These autosomal *'TDF'* genes may be responsible for the sexual dimorphisms preceding gonadal differentiation which are seen in marsupials (O *et al*,, 1988; discussed above).

Proof that *ZFY* does constitute the primary sex-determining gene could be obtained by the insertion of such a sequence by transgenic technology into the genome of an XX embryo to test for development of a male. Inserting an extra copy of the homologous *ZFX* sequence into an XX genome would test the 'dosage' model of sex determination (McLaren, 1988).

1.8 Repeated DNA sequences

In all species so far studied most of the Y chromosome consists of repeated DNA sequences. All mammalian DNAs contain a very large proportion of repeated sequences (Britten and Kohne, 1968). The proportion of the genome which is repeated can be estimated using the reassociation properties of DNA.

Double-stranded DNA can be rendered single-stranded by heating to a temperature sufficient to overcome the base stacking and pairing interactions that hold the strands together. This critical temperature, the Tm or melting temperature, is defined as the point where half of the strands have separated. When the temperature is decreased below the Tm, stable base pairing occurs and the single strands reassociate. The specificity of base pairing interactions (adenine to thymine, guanine to cytosine), allows only complementary sequences to base-pair for a distance sufficient to form a stable complex. Since the reassociation of complementary strands depends on random collisions, the rate of reassociation is a function of the concentration of complementary sequences while its extent depends on the time allowed for the reaction.

Reassociation is therefore expressed as a function of C_0t where C_0t=Concentration of nucleotides at zero time (moles/l) x time (sec) and has the units moles.seconds/litre.

Being bimolecular, the reaction displays second order kinetics described by $C/C_0 = 1/(1 + k.C_0t)$ (Britten *et al.*, 1974), where C_0 and C are concentrations of nucleotides initially and after time t, and k is the rate constant of the reaction whose value depends on the incubation conditions and the complexity (the number and types of sequences) of the DNA.

At a given total DNA concentration, highly repeated sequences have a higher concentration relative to less repeated sequences and will reassociate faster (i.e. at a lower C_0t value).

Renaturation analyses define three broad classes in the eukaryotic genome on the basis of their frequency (Davidson *et al.*, 1973). A small percentage of the genome consists of short, low complexity sequences of typically 5-250 bases, repeated up to 10^6 times and reassociating at very low C_0t values (Waring and Britten, 1966). A heterogeneous, moderately repeated class of related sequences with a range of repetition frequencies, 250-10,000 bases long, present at 20 to 10,000 copies renatures next. Unique sequences, present at one to a few copies per genome comprise the third class. This fraction probably encodes most structural genes (Lewin, 1975). In the DNA of a particular eukaryote each of these three classes has a characteristic C_0t value for the completion of its reassociation.

Some of the repeated sequences in a genome are interspersed amongst single copy sequences (Schmid and Deininger, 1975), or other, unrelated repeated sequences; others are in tandem arrays of a basic repeat unit , with minor base sequence variation (Brutlag, 1980). There are also interspersed long repeated sequence elements (LINEs, Singer, 1982), with no internal redundancy. Many roles have been proposed for this repeated fraction of the eukaryotic genome, including regulation of transcription (Britten and Davidson, 1969; Davidson *et al.*, 1973), production of evolutionary novelty (Britten and Davidson, 1971),encoding of regions in large RNAs to allow them to function as regulator/activators (Davidson and Britten 1979), or regulation of specific processing of large RNA transcripts (Davidson and Britten, 1979). The contrasting viewpoint is that most repeated DNA is selfish or parasitic (Doolittle and Sapienza, 1980; Orgel and Crick, 1980). Cavalier-Smith (1985) proposed that there is non-sequence-specific selection for increases or decreases in DNA content as a modulator of nuclear and cellular size. Dover (1986) suggested that

evolutionary change in the number of repeated DNAs seems to occur 'ignorant' of the sequence content of the segments involved, and proposed that such events may be an unavoidable by-product of the complexities of replication and recombination in eukaryotes. Whatever their role, if any, moderately repeated sequences are an integrated and established component of mammalian genomic DNA.

1.9 Y-chromosomal repeats

The Y chromosome appears to be a somewhat special case, consisting largely of repeated sequences, with probably very few unique sequences (genes). It also appears to be the least conserved among mammalian chromosomes (Lucchesi, 1978), in contrast with the X which is the most conserved (Ohno, 1967). Much of the Y chromosomal DNA of the mouse (Eicher and Washburn, 1986), and man (Goodfellow *et al.*, 1985), has been recently acquired and is highly repeated. Sequences from the human Y which have homology with autosomal sequences have been reported by Bishop *et al.* (1984), Heilig *et al.* (1984) and Koenig *et al.* (1985). These homologies are with sequences in different parts of the genome, which suggests the human Y may have arisen partly as a mosaic of autosomal sequences, perhaps constructed through a series of duplication and transposition events (Affara *et al.*, 1986).

It is possible that acquisitions from other chromosomes have been selected for simply to provide mass, perhaps to prevent the loss of the male-determining genes through a high frequency of non-disjunction events to which a very small Y chromosome may be vulnerable (Cooke, 1976). The accumulation of repeated sequences on the Y which may in time, lead to the evolution of Y-specific repeats may be selected for in a similar manner. Neutral mutations may also accumulate quickly on the Y chromosome because at least part of it is normally transmitted as a single haploid entity. The mouse Y chromosome also contains a large amount of murine retroviral related sequences, estimated by Phillips *et al.* (1982) to comprise up to 3% of the chromosome. It has been shown (Eicher *et al.*, 1982) that the majority of these sequences are contained within the central region of the Y and it was suggested that this region may be important in sperm motility.

Leroy *et al.* (1987) detected two pseudogene sequences on the human and mouse Y chromosomes, which detect testis-specific transcripts. The

original genes from which these pseudogenes arose are autosomally located, suggesting retroposition to the Y chromosome may be more active in the testis. The Y chromosome certainly represents a target area, with few essential sequences and it is possibly more accessible to colonization by retroviral transposition in the testis (Craig *et al.*, 1987). Bishop and Hatat (1987) isolated a testis-specific transcript from the mouse Y chromosome, which was not obviously of retroviral origin but was represented about 250 times and distributed at various positions along the length of the Y.

Current evidence indicates a very rapid evolution for the bulk of the mammalian Y chromosome. Viral transposition and adventitious colonization by members of repetitive sequence families is significant (Craig *et al.*, 1987). The heterochromatic long arm of the human Y chromosome is composed almost entirely of repeated DNA, and varies widely in length but not in sequence (McKay *et al.*, 1978). It may be that the Y chromosome and perhaps other polymorphic sites are evolutionary hot spots for repeated sequences. Alterations which may result from these and other Y-chromosomal rearrangements may modify the expression of critical genes in a temporal or quantitative manner and so could be important in speciation.

It has been proposed that those Y-specific genes responsible for male sex determination should become 'immune' from recombination and that surrounding sequences, isolated from the rest of the genome (Lucchesi, 1978), might duplicate themselves in the manner of selfish DNA (Doolittle and Sapienza, 1980). Proposed mechanisms for the production of such Y-specific repeats include saltation (Southern, 1970), and by sister chromatid exchanges (Smith, 1974). During replication these sequences might be expected to diverge rapidly, leading to the accumulation of species-specific repeats uniquely associated with the Y chromosome. Such species-specific repeats have been isolated from the mouse Y chromosome (Lamar and Palmer, 1984), and the human Y chromosome (Kunkel *et al.*, 1976). These sequences were isolated using methods based on the reassociation properties of DNA.

1.10 Identification of species'-specific Y-chromosomal repeats

Kunkel and his co-workers (1976) used 'exclusion hybridization' to isolate repeats specifically associated with the human Y chromosome. They obtained moderately repeated DNA by reassociation of radioactively-labelled human male genomic DNA to C_0t 46, followed by separation of the duplex repeat fraction from the single-stranded unique component. This fraction was subjected to repeated reassociation in the presence of a large excess of unlabelled female repetitive DNA. It was shown that radioactive sequences remaining single-stranded would reassociate with male but not female DNA. This male-specific group of 15-20 repeated sequence families, each repeated 300-600 times per Y chromosome, is termed it-Y DNA (iterated Y-specific DNA).

Using an alternative approach, Cooke (1976) digested human male DNA with the restriction endonuclease *Hae* III, which cuts on average every 256 bp. After gel electrophoresis he identified two bands of DNA that were not present in *Hae* III digests of female DNA. These same bands of 3.4 kb and 2.4 kb fragments were found in *Hae* III digests of human male satellite III DNA by Bostock and co-workers (1978). The 3.4 kb *Hae* III fragments are internally heterogeneous. There may be as many as 40 different families of male-specific sequences included within the set of 3.4 kb fragments, but only 3 of the possible 40 seem to be present in any particular fragment (Scott and Smith, 1982). The fragments appear tandemly repeated upon restriction because of the regularity of the restriction sites. The smaller *Hae* III fragments contain sequences with little variation. The two groups of fragments are unrelated, since they fail to hybridize with one another (Scott and Smith, 1982).

The 3.4 kb fragment has been estimated to comprise up to 40% of human Y-chromosomal DNA (Kunkel *et al.*, 1979). This repeat may have an autosomal origin as shown by its homology with autosomal sequences (Cooke and McKay, 1978). It has been suggested that transfer of a few autosomal copies to the Y chromosome was rapidly followed by amplification, mediated by unequal crossing over and/or duplication (Erickson and Goodfellow, 1984). These repeated sequences from the human Y chromosome are homologous with autosomal sequences in the great apes but not with their Y chromosomal sequences (Kunkel and

Smith, 1982), which also supports the view that these sequences have only recently been acquired by the human Y.

Both the 3.4 kb and 2.4 kb fragments are made up of Y-specific and non-Y-specific sequences. The non-Y-specific sequences include members of the *Alu* family repeats (Frommer *et al.*, 1984), 300,000 copies of which are interspersed throughout the genome (Houck *et al.*, 1979). The 3.4 kb *Hae* III fragment and other non-Y-specific repeat sequences combined make up 70 to 85% of the Y-chromosome.

Both the 3.4 kb fragment and it-Y DNA are restricted to the long, heterochromatic arm of the Y and play no part in sex determination (Kunkel *et al.*, 1977). The it-Y sequences are actually interspersed between the the highly repeated *Hae* III recognition sites of the 3.4 kb fragment, a rather unusual form of sequence interspersion (Davidson *et al.*, 1973), compared with the more common interspersion of intermediate repetitive sequences with unique sequences. The structural character of these fragments shows similarities with the scrambled repeats and some transposable elements in *Drosophila* (Rubin, 1983).

All of the Y-chromosomal 3.4 kb non-Y-specific sequences have a species-specific association with different specific autosomes and distinct domains (Burk *et al.*, 1985), and although many repeated sequence families often have homologues widely dispersed throughout the genome (Schmid and Deininger, 1975), the distribution of specific repeated DNA families and subsets of those families can be specific to chromosomes other than the Y. Satellite DNAs (highly repeated simple sequence DNA) such as the alphoid satellite sequences, are an example of such DNAs, usually having homologous sequences on most chromosomes but predominantly concentrated at centromeres and other heterochromatic sites (Miklos and John, 1979). Wolfe *et al.* (1985) isolated an alphoid centromeric repeat from the human Y. The sequence is 70% homologous to the alphoid consensus (Manuelidis, 1978) and 70% homologous to the X alphoid sequence.

The sex chromosomes, in particular the Y chromosome, often contain repeated sequences that are unique to that chromosome or are unique in lacking satellite sequences (Kurnit *et al.*, 1973; Kunkel *et al.*, 1976; Cooke and Hindley, 1979). These observations have led to the

hypothesis that it is the difference between the satellite DNAs of the sex chromosomes that is the first step in their evolutionary separation (Corneo, 1978).

The finding of sex-chromosome specific sequences in snakes has been used as evidence to support this hypothesis. Singh *et al.* (1980) have isolated a repeated sequence from snakes which is found predominantly on the W chromosome of the heterogametic WZ female in snakes with a well-defined sex chromosome system. This repeat is a minor component of the W satellite DNA of the Banded Krait and is designated Bkm DNA. Bkm DNA has been shown to hybridize to the Y chromosome of mice and is closely linked to the testis-determining gene (*Tdy*), in this species, since in sex-reversed male XX mice the Bkm sequence is co-translocated to the X with testis-determining loci in a minute fraction of Y-chromosomal material. The Bkm sequence in mice is also concentrated on the X chromosome and there are regional concentrations on two autosomes, one of which is on chromosome 17 near the major histocompatibility complex. The Bkm sequence is transcribed but it has been shown that the transcripts are detected in day 14 foetal gonads of both sexes (Erickson *et al.*, 1987). Although it is also found on the human Y chromosome (Arnemann *et al.*, 1986) the Bkm sequence cannot be involved with testis determination because it is not conserved in vertebrates (for example, it is not found in cattle and sheep: Miklos *et al.*, 1989).

It is unclear how such chromosome-specific organizations of sequences arise or how they are maintained. The function, if any, of such domains remains unknown. Raman and Nanda (1982) suggested that Y heterochromatin in mammals may be required for the production of a sex vesicle. The sex vesicle is a characteristic feature of early meiotic prophase in most species of mammals but does not occur in the Indian mongoose *Herpestes auropunctatus*. The significance of sex vesicle formation during meiosis is unclear but its appearance often signals the onset of condensation in X and Y chromosomes, followed by genetic inactivity of the X during spermatogenesis. The X and Y are also brought into intimate association by the sex vesicle for pairing and orientation of the sex chromosomes. The Y chromosome of the Indian mongoose lacks heterochromatin and undergoes meiosis as a translocation on an autosome. There may be a cause and effect relationship between the

deletion of heterochromatin from the translocated Y chromosome and the lack of sex vesicle formation.

In all of the species so far examined the Y chromosome has been found to contain repeated sequences specifically associated with that chromosome, although their functions, if any remain unknown. Repeated sequences specific to the Y chromosome may be used as a means of diagnosing the sex of an embryo.

1.11 Aims and Approach

The initial aim of this project was to isolate bovine Y-specific repeated sequences for use as probes in embryo sexing. Repeated sequences are more suitable as probes for embryo biopsies, simply because the higher copy number increases the sensitivity and speed of an assay. The bovine Y chromosome is very small (1.5×10^7 bp) and represents only about 1% of the genome in a complement of sixty chromosomes, so it was hoped that repeated sequences associated solely with the male genome might include, or provide access to, genes involved in male differentiation. Since other workers have successfully used the reassociation properties of DNA to isolate Y-specific sequences from other species the experimental approach used for this project incorporated the same basic principles. The obvious approaches, such as direct visualization of male-specific bands in restriction digests of genomic DNA had already been tried without success, and another technique involving the denaturation and renaturation of DNA in an agarose gel to give fine resolution of repeated sequences (Roninson, 1983), was also tried without success (Matthews, 1985).

Human Y-specific repeated and unique sequences have been used in the diagnosis of foetal sex where there is a risk of X-linked recessive disease (Vergnaud *et al.*, 1984). Human male embryos have been identified by the use of a commercially available DNA probe for Y-chromosomal DNA by means of in-situ hybridization to chorionic villus biopsy samples and amniocentesis samples (Burns *et al.*, 1985), and to cells biopsied from early cleavage to blastocyst stages of embryos (West *et al.*, 1987). Hybridization of Y-chromosomal probes to Southern blots of restriction endonuclease digests and to dot blots (Gosden *et al.*, 1984; Lau *et al.*, 1984) of DNA from chorionic villus biopsy samples has obviated the need to obtain chromosome spreads

and so improves the chances of being able to diagnose the sex of all the embryos.

Pre-determination of sex in livestock species has potentially great economic value, since all production of animals as food requires disproportionate numbers of one sex during some phase. Being able to obtain the optimal proportion of males to females would enable producers to benefit from sex-linked traits, such as milk production and sex-influenced traits, such as body composition and rate of weight gain. The speed and efficiency of selective breeding programs would also be increased by the elimination of superfluous males. Technology which would allow embryos to be sexed prior to implantation, when coupled with embryo transfer technology currently in use, would greatly increase the reproductive life and therefore the value of proven cows (Betteridge, 1989). The use of splitting techniques to increase the number of sexed embryos would further improve the economics of the situation. Over 115,000 unfrozen and 25,000 frozen bovine embryos were transferred in the United States and Canada in 1983 (Seidel, 1984). The huge amount of money and scientific resources which have been expended on trying to find a practical method of predetermining the sex of livestock embryos confirm the potential value of such technology.

Several methods have been used,sometimes commercially, with limited success (reviewed by Anderson, 1985). The procedure most commonly used is karyotyping a biopsy of cells from the embryo. The process is extremely tedious and time-consuming and though very accurate, in about one-third of cases a suitable spread of chromosomes cannot be obtained.

The other major approach has been the development of antibody to H-Y antigen, the male antigen defined by transplantation studies (Anderson,1987). This antibody is coupled to a second antibody with an attached fluorescent molecule. Male embryos are distinguished by their fluorescence. The main problem with the application of this approach commercially appears to be difficulty in training operators to make consistent assessments about basically subjective differences between embryos (Anderson, 1985). Williams (1986) reported that, by using a colorimetric test for the activity of the X-linked gene glucose-6-phosphate

dehydrogenase, he was able to correctly identify the sex of 64% of mouse embryos in his study. This method is based on the greater enzymatic activity of XX embryos compared with XY embryos, and has also been applied to cattle (Williams, 1986). The disadvantage of this method is that it requires a subjective assessment of the level of gene activity.

Very little is known about the bovine Y chromosome at the molecular level so a study of repeated sequences on this chromosome would represent an initial contribution in this area and perhaps extend the results of work on the Y chromosomes of other species.

1.12 Conclusion

Primary sex determination in mammals offers a (perhaps deceptively), simple system for the study of development in mammals, and as such has a great allure, as seen by the increasing interest in the field. Study of the mammalian Y chromosome also offers a relatively simple system for the study of the molecular organization of repeated sequences, and their possible functions. Many theories have been advanced concerning the roles apparently functionless repeated DNA sequences may fulfil, including that of regulation of gene expression (Davidson and Britten, 1979). Some repeats are now known to be needed for chromosome structure and integrity, for example those sequences found in the kinetochores and telomeres. Sequences associated with heterochromatin may fall into the same category.

The Y chromosome appears to be a special case in many ways, being composed in the species so far studied, of a very high proportion of repeated DNA, with only a limited number of functional genes, and with a very limited region participating in meiotic pairing (Cooke *et al.*, 1984; Simmler *et al.*, 1985). However, information gained from the understanding of one mammalian chromosome should provide generally applicable information about the evolution, propagation and maintenance of repeated sequences on all chromosomes.

CHAPTER TWO

A REPEATED SEQUENCE FROM THE Y CHROMOSOME OF CATTLE, SHEEP AND GOATS

CHAPTER TWO

A Repeated DNA Sequence from the Y Chromosome of Cattle, Sheep and Goats

2.1 Introduction

Since mammalian sexual dimorphism is due to the presence or absence of the Y chromosome (McCarrey and Abbott, 1979), the diagnosis of sex in embryos can be performed by determining whether a Y chromosome is present.

Despite intensive investigation over many years the mechanism by which the Y chromosome directs the development of the testes is still unknown. In most species a large part of the Y chromosome consists of repeated sequences. The acquisition of sequences from other chromosomes and the accumulation of Y-specific repeats may be selected for to prevent the loss of the male-determining gene(s) through a high frequency of non-disjunction events to which a very small Y chromosome may be vulnerable. It has been proposed (Lucchesi, 1978) that those Y-specific genes responsible for male sex determination should become 'immune' from recombination and that surrounding sequences, isolated from the rest of the genome, might duplicate themselves in the manner of selfish DNA (Doolittle and Sapienza, 1980). During replication these sequences might be expected to diverge rapidly, leading to the accumulation of species'-specific repeats uniquely associated with the Y chromosome.

Such species'-specific repeats have been isolated from the mouse Y chromosome (Lamar and Palmer, 1984) and the human Y chromosome (Kunkel *et al.*, 1976).

Human Y-specific repeated and unique sequences have been used in the diagnosis of the sex of foetuses where there is a risk of X-linked recessive disease (Vergnaud *et al.*, 1984). Repeated sequences specific to the Y chromosome are more desirable as probes since they are more easily detectable in assays using only small amounts of embryo biopsy material.

Pre-determination of sex in livestock species has potentially great economic value, since all animal food production requires disproportionate numbers of one sex during some phase. Being able to obtain the optimum ratio of males to females would enable producers to benefit from sex-linked traits, such as milk production, and sex-influenced traits, such as body composition and rate of weight gain. The speed and efficiency of selective breeding programs would also be increased by the elimination of superfluous males.

The aim then, was to isolate bovine Y-specific repeated sequences which could be used to develop probes for sexing preimplantation embryos, while increasing the understanding of the molecular organization of the Y chromosome. Y-specific sequences could also provide access to genes involved in testis determination.

2.2 Materials and Methods

2.21 Construction of a library by deletion enrichment

A library enriched for bovine Y-specific sequences was constructed using the deletion enrichment method of Lamar and Palmer (1984). This procedure enriches for sequences occurring uniquely or at greater abundance in one genome (for example the male genome) by 'deleting' those sequences shared with another genome (in this case the female genome).

Plasmid pUC11 (Vieira and Messing, 1982) was prepared by the alkaline lysis method (Birnboim and Doly, 1979), as modified by Ish-Horowicz and Burke (1981), and digested with *Bam* HI (New England Biolabs). All enzyme digestions were carried out according to the manufacturer's instructions. The *Bam* HI treated vector was then incubated with calf intestinal alkaline phosphatase (Boehringer) at 37°C for 1 h.

Female cattle liver DNA, purified by standard methods, was randomly sheared by passing through a French pressure cell five times at 20,000 psi. The size range of the fragments produced was determined by electrophoresis in a 1% agarose gel, to be 200-900 bp, with a mean size of approximately 450 bp.

Male cattle DNA was digested with *Sau* 3A1 (New England Biolabs) to generate fragments with 'sticky' ends complementary to those of the Bam HI restricted vector. Sheared female DNA and *Sau* 3A1 digested male DNA were mixed in a ratio of 100:1, denatured for 5 min at 105°C, and allowed to reanneal to a C_0t of 1320, in 200 µl of a buffer containing 2M ammonium sulphate, 50mM sodium phosphate buffer (pH 6.8) and 5mM EDTA, at 68°C for 24 h. The rate of reannealing is accelerated 50-fold in this buffer relative to the rate in 0.12M sodium phosphate buffer (D. Kohne, cited by Lamar and Palmer, 1984).

Hybrid double-stranded DNA was collected on a hydroxylapatite column (Bio-Rad), suspended in 0.12M phosphate buffer (pH 6.8) and maintained at 60°C. The column was washed 3 times with 0.12M phosphate buffer, then the reassociated DNA added in 5 ml of 0.12M phosphate buffer. The column was washed with 5 volumes of buffer and the double-stranded DNA was eluted with 0.6M phosphate buffer (pH 6.8). The double-stranded fraction was dialyzed against TE overnight to remove phosphate.

The renaturation of excess sheared female DNA with *Sau* 3A1-digested male DNA gives three types of double-stranded hybrids: (i) both strands derived from female DNA; (ii) one strand from female DNA, one strand from male DNA (including autosomal and X chromosomal DNA of the male as well as Y-encoded sequences homologous to autosomal and X-chromosomal DNA); (iii) both strands derived from male DNA. Only the male/male homodimers will have clonable ends.

The double-stranded fraction was ligated with the *Bam* HI digested pUC11 in a genomic DNA:plasmid DNA ratio of 20:1 (w/w). After overnight ligation at 14°C with 50 units of T4 DNA ligase (Pharmacia) in 66mM Tris-HCl (pH 7.5), 6.6mM $MgCl_2$, 1mM ATP, 100 µg/ml BSA and 10mM dithiothreitol, aliquots of competent *E.coli* JM83 cells (Vieira and Messing, 1982), prepared as described by Hanahan (1983), were transformed with 10 µl (0.4 µg) of ligated DNA. The transformed cells were plated on nitrocellulose filters (Schleicher & Schull) and grown overnight at 37°C on 2YT agar plates (Hanahan, 1983) containing 50 µg/ml ampicillin. After counting, the colonies were scraped off the filters into 10 ml of SOB media (Hanahan, 1983) containing 50 µg/ml ampicillin, diluted with 20% glycerol and dispersed; 20 µl aliquots of the

library were frozen in liquid N_2 and stored at -70°C. A high complexity library of an estimated 100,000 individual recombinants was obtained.

2.22 Southern blots

DNAs were cut with the appropriate enzyme (*Sau* 3A1 or *Bam* HI) and electrophoresed in a 1% agarose gel (Maniatis *et al.*, 1982) then transferred by a modified alkaline blotting procedure (Reed and Mann, 1985) to Zeta-Probe membrane (Bio-Rad). The membrane was pre-wet in hybridization solution: 2 x SSPE (Maniatis *et al.*, 1982), 7% SDS and 1% BSA (Church and Gilbert, 1984), then sandwiched between two sheets of Schleicher and Schull #589 filter paper and sealed in plastic. Hybridization was conducted overnight at 68°C. Washing was in 2 x SSC (Maniatis *et al.*, 1982), 0.1% SDS at room temperature for 15 min, followed by 0.5 x SSC, 1% SDS at 68°C for 30 min. Autoradiography was to X-Ray film (Fuji RX) with a DuPont Quanta III intensifying screen at -70°C.

2.23 DNA labelling

DNA (0.2 -0.4 µg) was labelled with [α-^{32}P]dCTP (10 mCi/ml, Amersham) by a modification of the nick-translation method of Rigby *et al.* (1977).

2.24 Dot Blots

Genomic DNAs were heated at 100°C for 10 min in 0.4M NaOH, 10mM EDTA, cooled to room temperature and serially diluted in 0.4M NaOH. The samples were applied to Zeta-Probe membrane using the Bio-dot (Bio-Rad) apparatus and hybridized and washed as described for Southern blots. For re-use membranes were stripped of probe by twice pouring on boiling TE and shaking till cool.

2.25 Northern Blots

RNA was isolated by Sandra Beaton from frozen tissues (Cathala *et al.*, 1983). The RNA was denatured by incubating at 50°C for 1 h in 1M deionized glyoxal, 50% DMSO and 10mM sodium phosphate (pH 7). After electrophoresis in 10 mM sodium phosphate the RNA was transferred to Zeta-Probe as described by Thomas (1983). The filters were washed in 2 x SSC twice for 15 min, then the RNA fixed to the membrane by u.v. irradiation for 5 min. The

filters were pre-washed by pouring on boiling 20mM Tris-HCl (pH 8) and shaking until cool. Overnight prehybridization was at 42°C in 5 x Denhardt's (Maniatis *et al.*, 1982), 0.1% SDS, 5 x SSPE and 50% deionized formamide with 0.5 mg/ml carrier DNA (autoclaved salmon sperm DNA; Sigma). After overnight hybridization at 42°C, the membrane was washed in 0.5 x SSC, 0.1% SDS for 15 min at room temperature, followed by a 30 min wash at room temperature in 0.2 x SSC, 0.1% SDS.

2.26 Library screening for Y-specific repeated sequences

A frozen aliquot was diluted and plated on 2YT agar with ampicillin to give approximately 5,000 colonies/plate. After overnight incubation at 37°C the colonies were transferred and fixed to nitrocellulose filters (Grunstein and Hogness, 1975). Pre-washing was in 0.1 x SSC, 0.5% SDS for 1 h at 65°C, followed by prehybridization overnight in hybridization solution (2.5 x SSPE, 5 x Denhardt's solution, 0.2% SDS) with 0.5 mg/ml carrier DNA. The filters were probed with pooled genomic DNA, labelled by nick translation, from 6 unrelated cows, in fresh hybridization solution at 68°C overnight. They were washed as described above for Southern blots and exposed to X-Ray film overnight at -70°C.

The developed film was aligned with the original filters and colonies not hybridizing to the labelled female DNA were isolated. Rapid colony lysates (Barnes, 1977) were prepared and the DNA electrophoresed in a 0.6% agarose gel so the plasmid size could be compared with that of the vector and background non-recombinants eliminated. Possible recombinants were selected and 'mini-preps' of plasmid DNAs were performed using a scaled-down version of the method used for large-scale plasmid DNA preparation (Maniatis *et al.*, 1982). This DNA was digested with *Sau* 3A1 to cut out the insert, electrophoresed in a 1% agarose gel, and transferred as described above to a Zeta-Probe membrane. This membrane was then hybridized with a probe of nick-translated DNA from 6 cows under the same conditions as for the colony filter hybridization. Recombinant plasmids which did not hybridize to the female probe were labelled by nick-translation and used to probe Southern blots of *Bam* HI digested genomic male and female cattle DNAs.

2.27 Sequencing

The fragment of interest was excised from pUC11 using *Eco* RI and *Hind* III, and purified by electrophoresis in 1% LMT agarose (Sigma). The gel slice containing the fragment was melted at 65°C, then the DNA extracted by an equal volume of phenol equilibrated with 10 x TAE. The upper phase was concentrated by butan-1-ol extraction of water, and after washing with IAC and ether (pre-washed with water), precipitated by addition of 0.25 volume of 10M ammonium acetate and two volumes of ethanol. The dried DNA pellet was dissolved in ligation buffer (as above) with T4 DNA ligase and ligated into the *Eco* RI and *Hind* III sites of the vector pTZ18U or pTZ19U (Bio-Rad). The ligation mix was transformed into competent *E. coli* JPA101 cells and plated onto M9 plates (Maniatis *et al.*, 1982).

Competent *E. coli* JPA101 cells (Yanisch-Perron *et al.*, 1985) were prepared by inoculating 50 ml of LB media containing 0.2% (w/v) glucose with frozen stock, then incubating at 37°C with vigorous shaking to A_{600}=0.6. The culture was centrifuged at 6000 g for 5 min, the cells resuspended in 25 ml of cold 50 mM $CaCl_2$ and allowed to stand on ice for 20 minutes. Then the cells were again pelleted and resuspended in 50 mM $CaCl_2$ containing 14% (v/v) glycerol. Aliquots of 100 µl were stored at -70°C.

Single-stranded DNA templates were prepared by incubating transformed JPA101 in 2 ml of 2YT with 50 µg/ml ampicillin, with approximately 10^8 pfu of the helper phage M13K07. This helper phage uses the f1 origin of replication in pTZ18U and pTZ19U to generate single-stranded DNA copies of the plasmid (Vieira and Messing, 1987). After about one hour at 37°C with shaking, kanamycin was added to 70 µg/ml to kill any cells which had not been super-infected with M13K07, and the incubation was continued overnight.

1.5 ml of the culture was centrifuged at 13,000 rpm in an Eppendorf centrifuge for three minutes and 1.2 ml of the supernatant was incubated at room temperature for 5 min with 20 µg/ml with RNAse A (Boehringer). PEG 6000 (20% in 2.5 M NaCl) was added to precipitate the phage and allowed to stand at room temperature for 15 minutes, then the phage particles were recovered by centrifugation at 13,000 rpm for 5 min in an Eppendorf centrifuge.

The supernatant was discarded, then the pellets were recentrifuged and any remaining PEG was removed. The phage pellets were resuspended in 100 µl of TE and the ssDNA extracted by vigorous vortexing with an equal volume of phenol. After centrifugation for 15 min at 13,000 rpm in an Eppendorf centrifuge the upper phase was extracted by vortexing with an equal volume of IAC, followed by centrifugation as before. The upper phase was collected and the DNA precipitated with ethanol. The ssDNA was dissolved in 50 µl of TE and stored at -20°C.

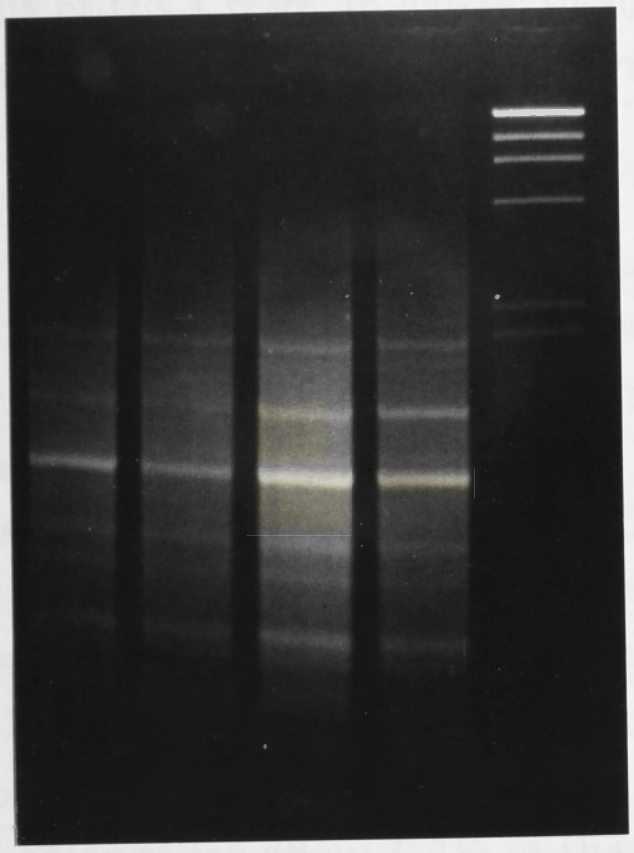
The nucleotide sequence was determined from both strands using separate ss templates by the dideoxy method (Sanger *et al.*, 1977), using [α-^{35}S]dATP and Sequenase (USB) according to the instructions of the manufacturer, with 7-deaza dGTP replacing dGTP (Mizusawa *et al.*, 1986). Electrophoresis was in 5% (w/v) acrylamide:bisacrylamide (19:1) gels containing 7 M urea in 1.5 x TBE, using the Bio-Rad Sequi-Gen apparatus with 0.4 mm spacers. Sequencing gels were fixed for 20 min in 10% methanol and 10% acetic acid before drying and exposure to Fuji RX or Kodak XAR X-ray film.

The autoradiographs were read with digitizing hardware and software (Gene-master Version 1; Bio-Rad) and analysed using sequence comparison software provided by the Gene-Master system.

2.3 Results

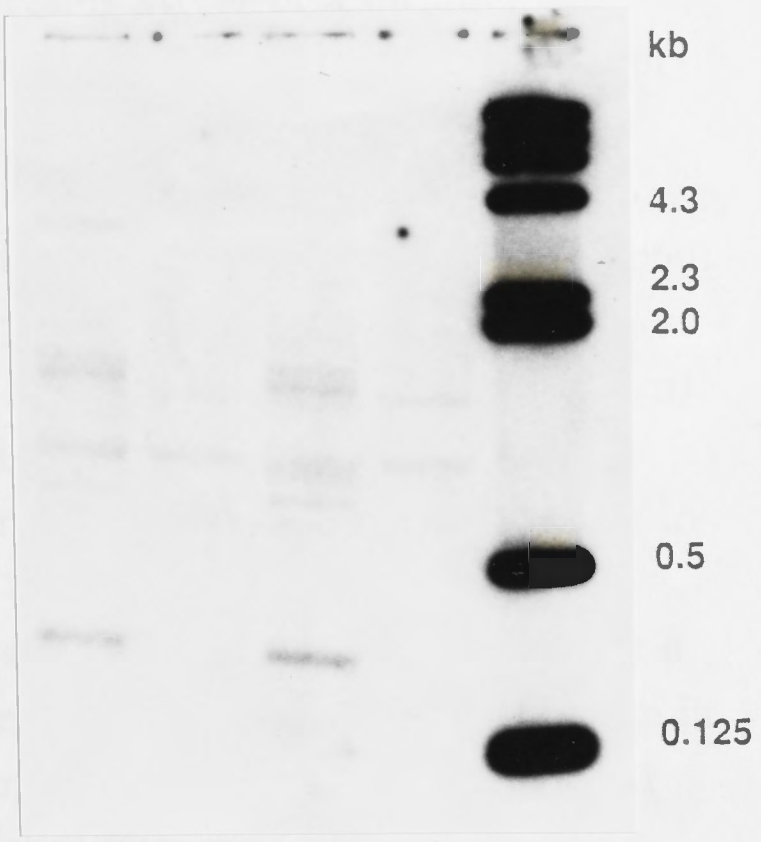
2.31 A Y-associated repeated sequence

Under the conditions used all of the recombinants screened, except one, hybridized to both male and female DNA. This plasmid contains an inserted *Sau* 3A1 fragment of bovine DNA which, as seen in Figure 2.1, appears to be male-specific. The fragment hybridizes to a single *Sau* 3AI fragment of genomic bull DNA of about 300 bp, which is the cognate band corresponding to the fragment cloned. After a long exposure (2 weeks), several other faint bands are also seen, at 600, 700, 900 and 1000 bp, in the male. There are also several bands at about 1.6, 1.7 and 1.8 kb. In the female corresponding faint bands of hybridization are seen at 1 and 1.6 kb, but these bands are seen only after a very long exposure to fast film, on a Southern with large amounts of



1 2 3 4 5

2.1A



1 2 3 4 5

2.1B

Figure 2.1. 5 µg of male (lanes 1 and 3) or female (lanes 2 and 4) cattle DNA was digested with *Sau* 3A1 and electrophoresed in a 1% agarose gel, (**2.1A**), then transferred by alkaline blotting to Zeta-Probe membrane and probed with nick-translated BRY.1 (**2.1B**). Lane 5 contains endlabelled λ *Hind* III markers.

genomic DNA. The sequence also hybridizes to several *Bam* HI fragments of male DNA (see below).

The sequence has been designated BRY.1, denoting the first isolated Bovine Repeated sequence associated with the Y chromosome. Hybridization analysis of DNA from 78 cattle, including representatives of a number of breeds, both unrelated, and of known pedigree, confirmed the Y-specificity of BRY.1 (data not shown).

BRY.1 is quantitatively polymorphic, varying in copy number from about 5 to 50 copies between individual males. Copy number was estimated using a known amount of the BRY.1 insert as a copy number control (data not shown). Figure 2.2 shows an autoradiograph of a dot blot of DNAs from male and female cattle, probed with BRY.1. The sequence is present at varying levels in all the males examined but is not seen in any of the females at this level of stringency. The sequence is present in females as seen in Figure 2.1, but at a much lower level, about one to two copies per genome. If the Y chromosome is 0.5% of a haploid genome containing 3×10^9 bp of DNA (Lewin, 1974), BRY.1 would account for about 0.1% of the cattle Y chromosome.

BRY.1 was used to probe Northern blots of poly(A)$^+$ RNAs from various bovine tissues and there was no evidence of transcription in any of the tissues examined.

2.32 Sequence of BRY.1

BRY.1 was subcloned into the pTZ18U and pTZ19U plasmids and sequenced in both directions as described in Materials and Methods. The fragment is 306 bp long (Appendix 1) and contains short blocks of A and T. There is no significant homology between BRY.1 and any sequence on the Genbank files (Genbank Release 48, 1987). The sequence contains no identifiable open reading frames.

2.33 BRY.1 in other species

A Southern blot of *Bam* HI digested cattle, sheep, goat, deer and pig DNA from both sexes was probed with nick-translated BRY.1. As can be seen in Figure 2.3, BRY.1 hybridizes to several *Bam* HI fragments of cattle male

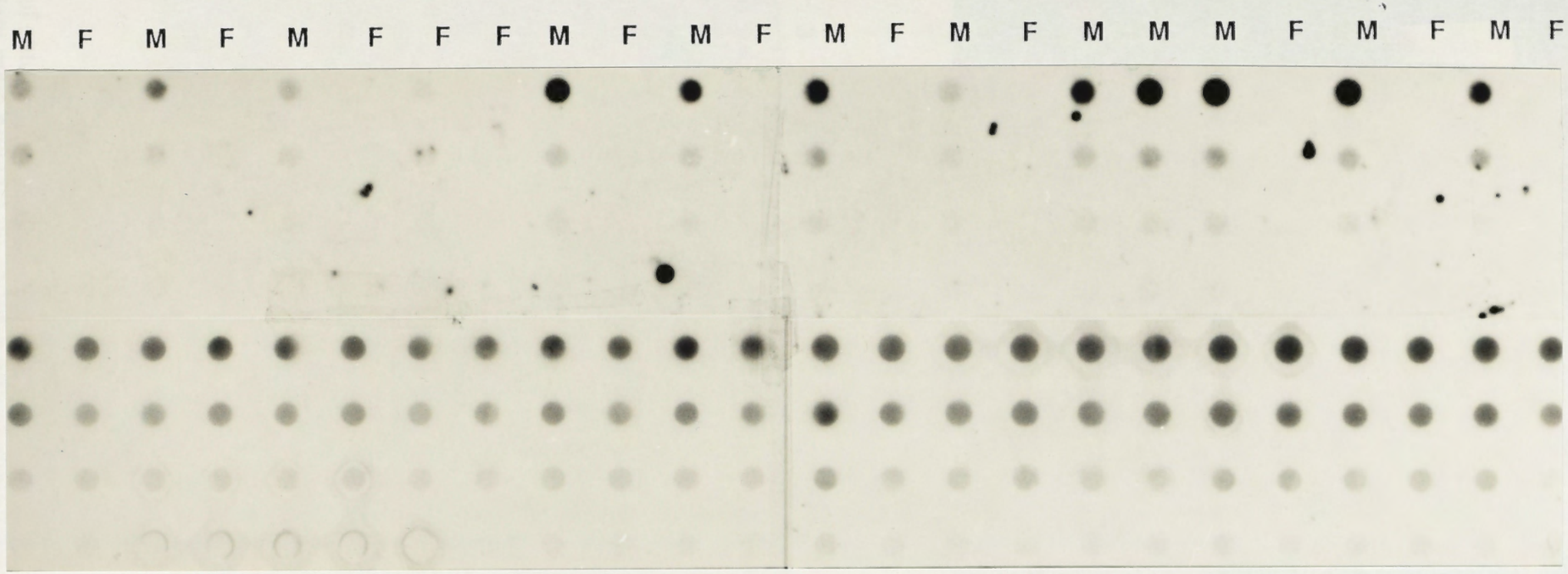
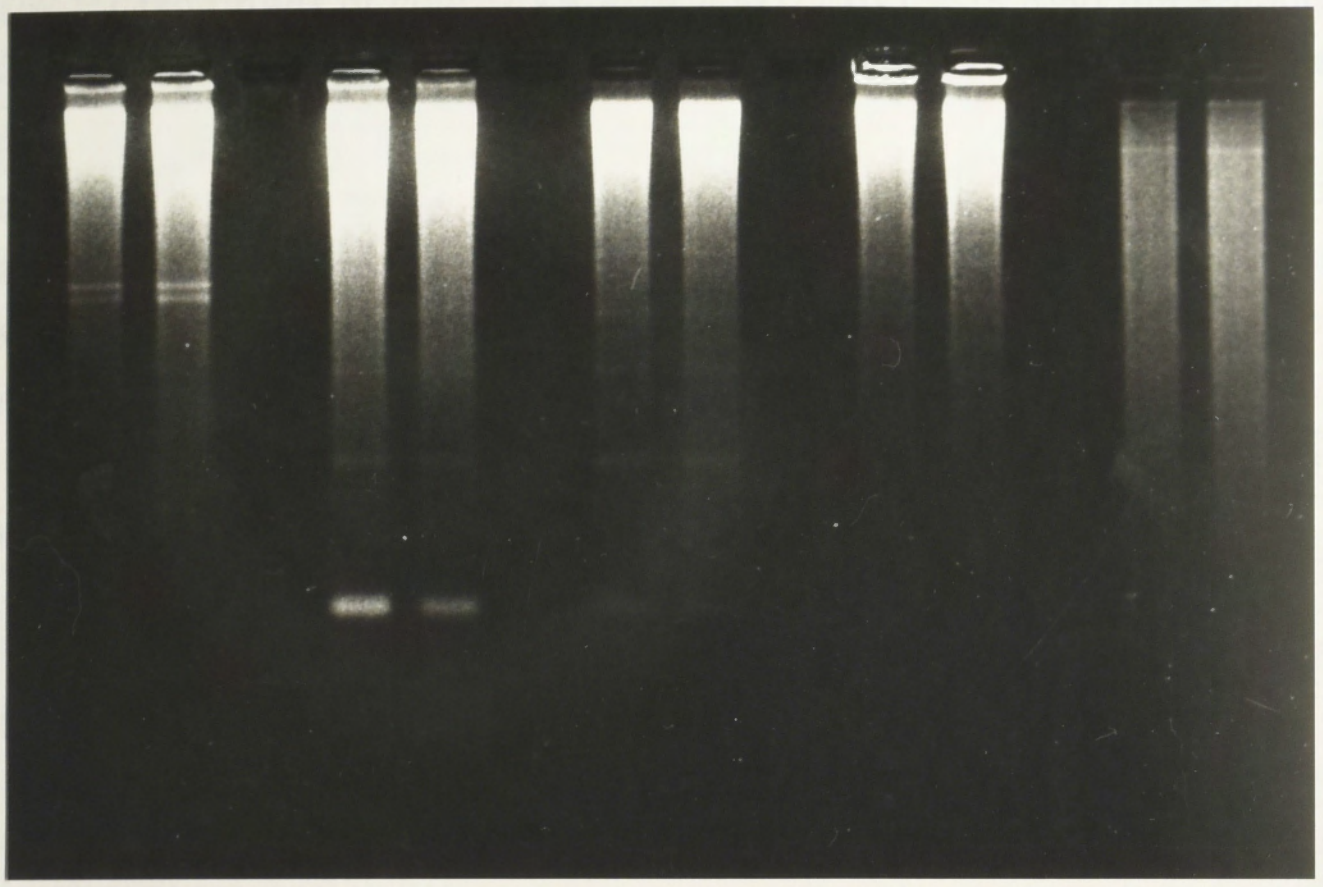
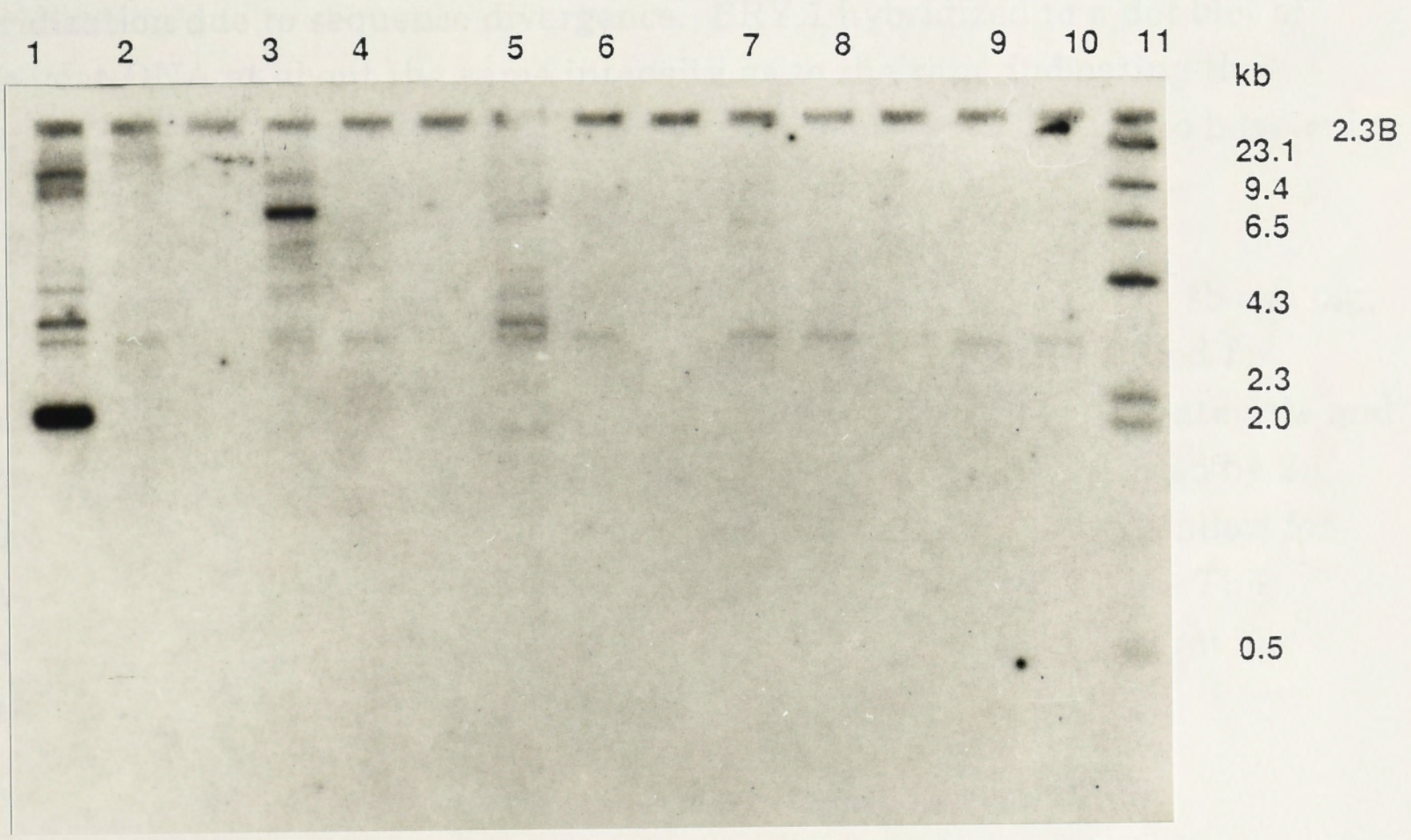


Figure 2.2. Alkali treated DNA samples from 24 cattle (Male and Female) were applied to Zeta-Probe membrane using a Bio-Rad dot blot apparatus. This membrane was probed: **a)** using BRY.1 labelled by nick-translation. The dot blot shows this sequence is present at varying levels in all the males examined but is absent from all the females.

To verify the apparent quantitative difference between males the membrane was stripped, checked for the absence of radioactivity and reprobed: **b)** with a radioactive nick-translated clone of the autosomal cattle satellite 1, which is present at the same level in all individuals (Matthaei and Reed, 1986). The result confirms the DNA samples were loaded with reasonable consistency and that BRY.1 is quantitatively polymorphic between males.

2.3A





2.3B

Figure 2.3. 2 μg of cattle male and female DNA (lanes 1 and 2), sheep male and female (lanes 3 and 4), goat male and female (lanes 5 and 6), deer male and female (lanes 7 and 8) and pig male and female (lanes 9 and 10) DNAs were digested with *Bam* HI and electrophoresed overnight at 12 volts in a 1% agarose gel, (**2.3A**). The DNAs were transferred by alkaline blotting to Zeta-Probe membrane and probed with nick-translated BRY.1, (**2.3B**). Lane 11 contains endlabelled λ *Hind* III markers.

DNA: 2.2, 3.2, 3.7, 4.2, 4.4, 9.4 and 14 kb in size. BRY.1 hybridizes faintly to a *Bam* HI fragment of cow DNA at 3.2 kb. BRY.1 is also male-specific in sheep, hybridizing to several bands in the ram DNA, with very little hybridization to the female. The homologous ovine sequence is found in *Bam* HI restriction fragments of 1, 3.2 and 7 kb, with hybridization to minor bands from 4.2 to 5 kb. There is one copy of the sequence in female DNA in a *Bam* HI fragment of 3.2 kb as for the cow. BRY.1 is also male-specific in goats, another important domestic species, hybridizing to *Bam* HI fragments of 1.1, 2, 3.2 and 3.7 kb,and to minor bands from 3.7 kb to 4.2 kb. The female of this species also appears to have a copy of the sequence in a 3.2 kb *Bam* HI fragment.

BRY.1 hybridizes to a *Bam* HI fragment of about 3.2 kb in both sexes of fallow deer (*Dama dama*) and to a 3 kb fragment in both sexes of the pig.

The copy number of the BRY.1 homologue in sheep was estimated, as described above, at about 10 copies per male. The copy number may be lower in sheep as compared with cattle, but this difference may be due to weaker hybridization due to sequence divergence. BRY.1 hybridized to a dot blot of male goat DNA at about the same intensity as to the ram, indicating the presence of about 10 copies per genome. Both sexes of deer appear to have only one or two copies of the BRY.1 homologue.

BRY.1 was used to probe a dot blot of male and female cattle, sheep, pig, human, mouse and rat DNAs (data not shown). The conditions used for washing this dot blot were less stringent than those described in Materials and Methods: 2 x SSC, 0.1% SDS at room temperature for 15 min, followed by 2 x SSC, 1% SDS at 68°C for 30 min. The lower stringency was used to allow for sequence divergence between species leading to less stable hybrids. This experiment showed that there is no homology to BRY.1 in human, rat or mouse.

2.4 Discussion

The complex screening procedure used was necessary because the deletion enrichment method depends on the two genomes used being identical, except for the presence of Y-specific sequences in the male genome which are

'deleted' from the female genome. This requirement is not a problem when inbred strains of mice are used but unfortunately such inbred cattle lines are not available, so that the DNAs contain varying amounts of many repeated sequences and the library is enriched for quantitative polymorphisms from the X and autosomes.

The deletion enrichment method may be useful when polymorphic sequences are desired, for example, a library constructed in this way would very quickly yield a bank of quantitatively polymorphic sequences linked to loci which could be correlated with specific traits.

It should be noted that because BRY.1 has been isolated by enrichment cloning the sequence obtained for this representative clone may be atypical of other BRY.1 elements. It is possible that the two *Sau* 3A1 fragments which reannealed were not perfect complements and that mismatch repair has occurred after transformation into *E. coli*.

This is the first repeated sequence associated specifically with the Y chromosome of cattle, sheep and goats which has been isolated. BRY.1 is also the first Y-specific repeated sequence to be described which is conserved on the Y chromosome between species. This remarkable conservation, across 15 to 20 million years (Novacek, 1982), may indicate that this sequence is from a highly conserved region of the Y chromosome.

BRY.1 does not appear to be tandemly repeated on the Y chromosome of cattle, sheep and goats, because no 'ladder' pattern is seen when a Southern blot of male DNA is probed with BRY.1. A ladder pattern is typical for tandem repeats which may have lost one or two restriction sites, leading to multimers of the fragment size on hybridization of the probe to a DNA digest (e.g. Cooke *et al.*, 1982). It is possible that a different enzyme may reveal such a pattern. BRY.1 is probably interspersed on the Y chromosomes of cattle, sheep and goats with other sequences since it hybridizes to restriction fragments of different sizes in the DNA of males from each of the three species (Figure 2.3). It is unclear whether these changes represent altered restriction sites or insertion/deletion events. Although the differences in restriction patterns among the species may be the result of sequence divergence, the stringent

wash conditions ensuring extensive sequence homology between the cattle sequence probe and the restriction fragments of sheep and goat DNA suggest the changes may be due to insertion or deletion events.

Leonard *et al.* (1987) reported that they had cloned a bovine Y chromosome-specific sequence which is repeated 2,000 to 2,500 times; however this sequence is Y-specific only in cattle. Lamar and Palmer (1984) used the deletion enrichment method to isolate a mouse Y-specific sequence which proved to be strain-specific. The human Y-specific repeats so far isolated (Cooke, 1976) are also found on the Y chromosome only in man, although they have homologues (at low stringency) on the autosomes of other primates. These homologues are not found on the Y chromosome in these primates (Kunkel and Smith, 1982). The human evolutionary line has been separate from those of gorilla and chimpanzee for the same period that the Bovinae (cattle) and Caprinae (sheep and goats) have been separated (Novacek, 1982).

The isolation of sequences associated mainly with the Y chromosome, which have been conserved across the evolutionary distance of 15 to 20 million years since the Caprinae, and the Bovinae diverged, was therefore unexpected. In some respects this result seems to be in conflict with the original hypothesis that Y-specific repeats would arise on the cattle Y chromosome due to isolation and rapid divergence from the rest of the genome, this isolation being due to the need to protect the male-determining gene(s) from recombination.

The finding that human Y-specific sequences are absent from the Y chromosome of other primates but present on the autosomes in those species has been used as evidence for the recent evolution of the human Y (Kunkel and Smith, 1982). The most likely explanation for the absence of these sequences on the Y chromosome of other primates is that in the primate progenitor these sequences were autosomal and only the human Y chromosome has acquired them, the autosomal representatives being lost in the human. The presence of BRY.1 almost exclusively on the Y chromosome of cattle, sheep and goats suggests that it was indeed present on the ancestral Bovidae Y chromosome, since it is unlikely to have been acquired twice during the course of evolution. Maintenance of a repeat over this length of time suggests that the Y chromosome in this group has remained remarkably stable over a long period,

especially in comparison with that of the primates. There is strong cytological evidence (Wurster and Benirschke, 1968) that the chief mechanism of karyotype evolution in the Bovidae is Robertsonian translocation (Robertson, 1916), and artiodactyls in general are unusual in having undergone few evolutionary changes in fundamental chromosome number. C-banding has shown a considerable degree of homology between the sex chromosomes of cattle, sheep and goats (Evans *et al.*, 1973), reflecting a common evolutionary origin, though the Y chromosome is very small in sheep and goats in comparison to that in cattle (Schnedl and Czaker, 1974). The cattle Y has centromeric heterochromatin extending into the arms, while the goat and sheep Y chromosomes lack this heterochromatin (Evans *et al.*, 1973). This extra DNA may include the Y-specific repeats reported by Leonard *et al.* (1987) which are found only in cattle.

The presence of BRY.1 in equivalent amounts in both sexes in deer would seem to suggest that either the progenitor artiodactyl leading to the Bovidae line lost these sequences from other chromosomes or that the sequences related to BRY.1 have been amplified on the Bovidae Y chromosome since the divergence of the Bovidae and the Cervidae 30 million years ago (Goodman *et al.*, 1982). It is interesting to note that the sequence is also present in the pig (Family Suidae), which has been separate from the Cervidae and Bovidae for 56 million years. A later amplification of this sequence seems more likely than the loss of BRY.1 from the Y chromosome in both the cervids and the suids. A phylogeny of the order Artiodactyla is shown in Figure 2.4.

The fact that this sequence has been so well conserved may indicate it is from an area of the Y chromosome which is prevented by some mechanism from rapid change, perhaps due to an association with element(s) crucial for male sex determination. Selective pressure for the maintenance of segments of repeated sequences may arise when such a sequence is linked to loci subject to positive selection (Bouchard, 1982). In this way a repeated sequence may be able to 'hitch-hike', and thus maintain or increase its numbers in the genome. It seems reasonable to suggest that repeated sequences resulting from a recent replication event in a lineage undergoing speciation episodes, some of which will be either temporarily abundant or linked to genes, might have an increased chance of fixation during the formation of a new species' genome.

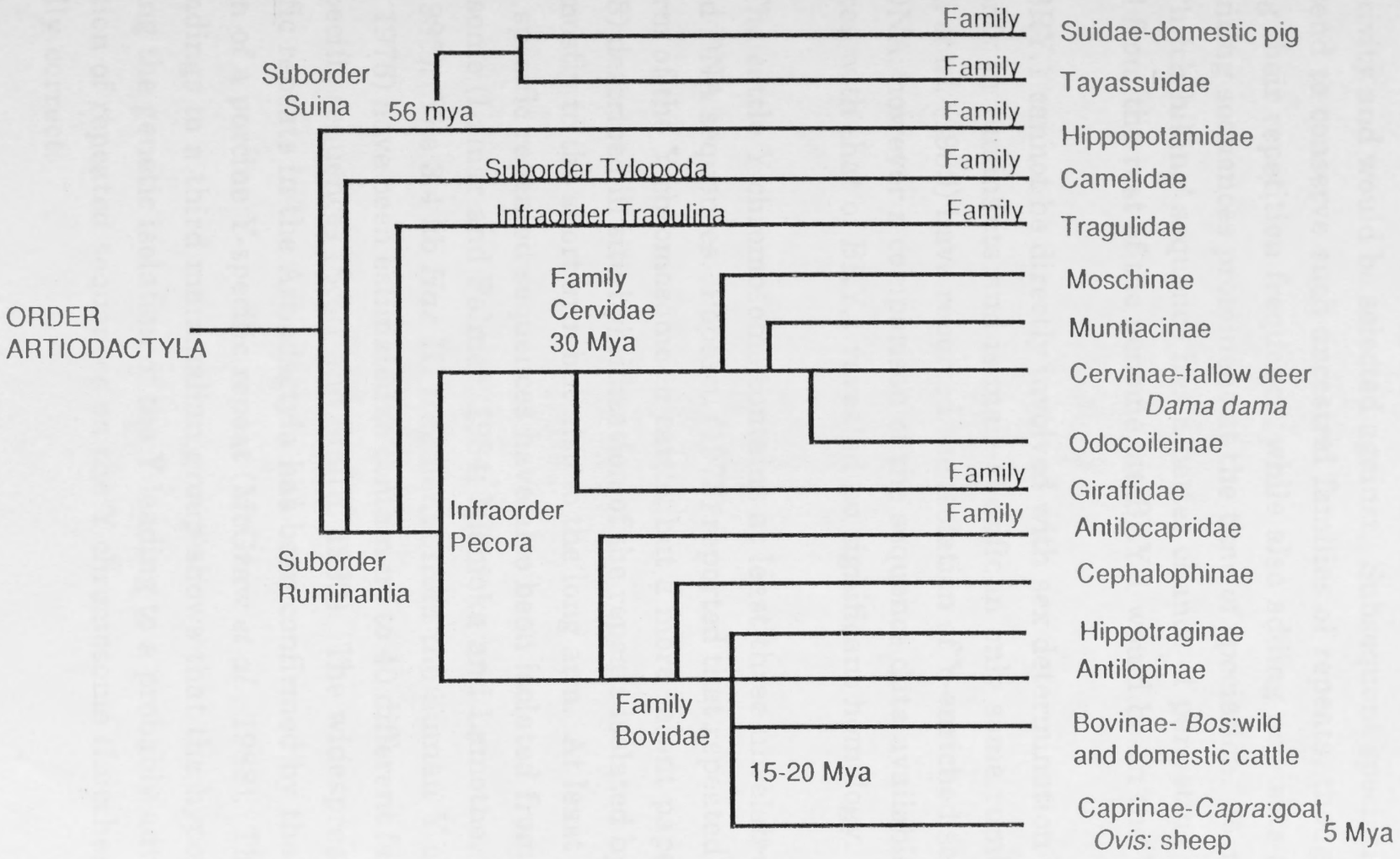


Figure 2.4. Diagram of the classification of the Artiodactyla, from Eisenberg (1981). The figures refer to the length of time (Million years ago) since the various groups diverged.

Once a set of families has been established in a lineage, the families should tend to persist for long periods and through many further bifurcations of the lineage. Later deletion of such sequences would be particularly difficult if they had successfully inserted into the immediate domains of genetically functional regions, since random deletions would remove adjacent regions required for gene activity and would be selected against. Subsequent speciation events would tend to conserve such ancestral families of repeats, though perhaps lowering their repetition frequency, while also adding new sets of families representing sequences prominent at the time of speciation. It may be that such a 'hitch-hiking' sequence has a better chance of persisting when it is isolated from the rest of the genome, as BRY.1 would be on the Y chromosome.

BRY.1 cannot be directly involved with sex determination because it is found only in ruminants and is male-specific in only some ruminant species. Bondioli *et al.* (1989) have reported the isolation of Y-enriched sequences from cattle DNA; however a comparison of the sequence data available for these sequences with that of BRY.1 revealed no significant homology.

The cattle Y chromosome contains at least three unrelated families of repeated DNA sequences. Popescu (1973) reported that repeated DNA is on the short arm of the Y chromosome in cattle, but a more recent paper (Popescu *et al.*, 1988) described in situ hybridization of the repeats isolated by Leonard *et al.* (1987) mostly to the short arm but also to the long arm. At least three families of male specific repeated sequences have also been isolated from the mouse Y chromosome (Lamar and Palmer, 1984; Nishioka and Lamothe, 1986; Bishop *et al.*, 1985). The 3.4 kb *Hae* III fragments from the human Y chromosome (Cooke, 1976) have been estimated to contain up to 40 different families of male-specific sequences (Scott and Smith, 1982). The widespread occurrence of Y-specific repeats in the Artiodactyla has been confirmed by the recent isolation of a porcine Y-specific repeat (McGraw *et al.*, 1988). The extension of such findings to a third mammalian group shows that the hypothesis regarding the genetic isolation of the Y leading to a probably adventitious replication of repeated sequences on the Y chromosome (Lucchesi, 1978) is generally correct.

Further study of the sequences surrounding BRY.1 on the sheep and cattle Y chromosomes was undertaken so that the length of the repeat unit and the relationship of this sequence to others on the Y chromosome could be better understood. Such further characterization may provide information about the significance of such conserved repeated sequences.

CHAPTER THREE

SEQUENCES FROM A FAMILY OF BOVINE Y-CHROMOSOMAL REPEATS

CHAPTER THREE

Sequences from a Family of Bovine Y-Chromosomal Repeats

3.1 Introduction

The isolation of BRY.1 from the bovine Y chromosome raised several questions concerning the organization of this repeat on the chromosome. BRY.1 is a *Sau* 3A1 fragment from the cattle Y chromosome which has a Y-associated homologue in sheep and goats. The deer and pig homologues are not found in higher numbers on the Y chromosome, but are present at close to single copy in both sexes.

The original hypothesis had been that a repeated sequence conserved between species on the Y chromosome may have been protected from recombination due to its proximity to unique sequences, perhaps even the gene(s) responsible for primary male sex determination.

BRY.1 is not tandemly repeated but occurs instead in several *Bam* HI fragments from the Y chromosome in cattle, sheep and goats. The presence of this sequence in restriction fragments of differing sizes and in varying copy numbers in each species make it unlikely that it has been conserved solely because of its proximity to coding regions. It may be more reasonable to infer positive selection for the retention of a cluster of tandem repeats, which may be adjacent to regions important for survival such as the sex-determining gene(s).

Repeated sequences, as a group, diverge more rapidly than do single copy sequences (Rice, 1972; Deininger and Schmid, 1979). The isolation of a repeated, presumably non-coding sequence which has been conserved across an evolutionary distance of 15-20 million years (Novacek, 1982) and which has remained confined to a single chromosome therefore poses special problems in understanding the mechanisms of its evolution and maintenance.

One proposed method for the evolution of repeated sequences (Smith, 1976) is an unequal crossover type of mechanism where the continuous sampling or error of genetic drift is thought to generate repeated DNA,

maintain sequence homogeneity within a species and generate differences between species . The obvious difficulty with this type of model to explain the evolution of BRY.1 is that the sequence has not diverged a great deal in the various species, thus presumably has not been subject to much genetic drift since its original transfer to the Y chromosome of a progenitor ruminant and a possible subsequent amplification event.

The variation in sequence number observed, without a concurrent alteration in the sequence itself, suggests the amplification may have been by a saltatory rather than a continuous process (McKay *et al.*, 1978), that is, in a few large steps rather than slowly and continuously. As an example, McKay and co-workers used the long arm of the human Y chromosome which varies widely in length but not in sequence, and suggested the Y chromosome may be an evolutionary 'hot spot' for repeated DNA sequences.

Perhaps the Y chromosome of Bovidae species is also a 'hot spot' for recombination of repeated sequences, because the fact that the sizes of the restriction fragments containing BRY.1 differ between species suggests that different sequences (or at least sequences of different length) may be interspersed between each copy of BRY.1 on the Y chromosome of each species. The differing sizes may also be due simply to point mutations leading to the loss or addition of restriction enzyme sites. There was a need to determine the repeat unit length, since the 306 bp *Sau* 3A1 fragment isolated may be part of a much longer sequence which may even be transcribed thus providing some clues as to the reason for its conservation.

In order to learn more about the organisation of this repeat on the chromosome and its position relative to other sequences, BRY.1 was used as a probe to isolate a larger fragment of DNA from the bovine Y chromosome.

3.2 Materials and Methods

3.21 Phage library screening

An amplified λEMBL3A library (constructed in this laboratory by Francis Stewart) of *Sau* 3A1 partially digested male cattle genomic DNA, cloned into the *Bam* HI-restricted vector, was screened for recombinant phage containing sequences homologous to BRY.1. The library was four

years old and had been stored in SM/chloroform at 4°C (Maniatis *et al.*, 1982). The titre was established to be 10^6 pfu/ml by plating on *E. coli* LE392 (Murray *et al.*, 1977), grown by the method of Maniatis *et al.* (1982).

An aliquot was diluted in SM to give approximately 40,000 plaques per 150 mm plate, adsorbed to 100 µl of Mg^{2+}-treated LE392 cells for 30 min at room temperature, then added to 3.5 ml of top agarose and plated on NZCYM plates (Maniatis *et al.*, 1982). A total of 240,000 plaques were screened, equivalent to just over a haploid genome equivalent if an average insert of 15 kb/phage is assumed.

When the plaques were visible but not confluent (9-10 h), the plates were placed at 4°C for 2 h to harden the top agarose. Nitrocellulose filter (Schleicher and Schull) lifts were made (Benton and Davis, 1977) and the DNA fixed and baked to the filters. The filters were pre-washed in 4.5 x PE, 0.2% SDS at 42°C for 30 min, then prehybridized (2 filters/bag) in 20 ml of 5 x SSPE, 0.4% SDS, 0.2% BLOTTO (Johnson, 1984) and 0.2 mg/ml carrier DNA at 68°C overnight. A fresh 20 ml of hybridization solution containing pG.BRY.1, labelled by nick-translation with [α-^{32}P]dCTP, was allowed to hybridize overnight at 68°C.

The filters were washed at room temperature for 15 min in 2 x SSC, 0.1% SDS, then at 68°C for 30 min in 2 x SSC, 0.2% SDS. The filters were wrapped in Saran wrap and exposed to Fuji RX X-ray film for 2 days at -70°C with an intensifying screen (DuPont Quanta III).

Putatively positive plaques were identified by aligning the film with the filters and the filters with the plates. A plug of top agarose and agar from that area of the plate was taken into 500 µl SM and left for 4 h at 4°C in 500 µl SM to allow the bacteriophage particles to diffuse out of the agar. It was assumed each plug contained 10^6-10^7 pfu and the phage stock was diluted to give 40 pfu per plate.

The screening procedure was repeated until, after the fourth round, all plaques on each plate aligned with a positive signal on the X-ray.

3.22 Isolation of phage DNA

Phage DNA was isolated from the positive plaques by taking the plug of top agarose into 500 µl SM, then diluting 1 µl of this stock into 249 µl SM. One-fifth of this dilution was adsorbed to 50 µl LE392 cells for 30 min at room temperature. The cells were then added to 40 ml of NZCYM and grown with shaking at 37°C overnight. Five drops of chloroform, RNAse A to 2 µg/ml and DNAse I (Boehringer) to 1 µg/ml were then added to the culture and incubation continued for 60 min at 37°C.

The culture was centrifuged at 16,000 x g for 10 min, then the supernatant was removed and centrifuged at 27,000 rpm for 1 h at 4°C in a SW27 rotor in a Beckman centrifuge. The supernatant was discarded and the phage pellet dissolved in 2 ml of SM buffer. EDTA (pH 8) was added to 20mM with SDS to 0.5%, and Proteinase K (Boehringer) at 50 µg/ml and the DNA incubated at 60°C for 1 h. The phage DNA was extracted twice with phenol equilibrated with 10 x TAE, and once with IAC.

The phage DNA solution was concentrated to 0.8 ml, loaded on a 1M NaCl/TE cushion and centrifuged at 58,000 rpm in a Ti-80 rotor in a Beckman L8-80 centrifuge for 2 h at 20°C. The DNA pellet was dissolved in 0.5 ml of TE buffer and the concentration estimated by spectrophotometer reading (Maniatis *et al.*, 1982). Phage DNA was stored over chloroform at 4°C.

3.23 Restriction mapping of phage using end-labelling

1 µg of phage DNA was heated at 65°C for 5 min, then chilled and 5 units of Klenow fragment (Pharmacia) added with 0.2mM dGTP, 0.2mM ddATP (Promega) and 1 µl [α-^{32}P]dCTP. The reaction was allowed to stand at room temperature for 15 min. The inclusion of dGTP allows [α-^{32}P]dCTP labelling of the right *cos* arm of the phage, while ddATP prevents labelling of the left arm.

The phage DNA was then partially digested with the appropriate restriction enzyme (New England Biolabs) under conditions recommended by the supplier for 5, 10 or 15 min. The labelled, restricted phage DNA was then electrophoresed in a 0.4% agarose gel, with end-labelled λ *Hind* III and λ *Acc* I markers (Maniatis *et al.*, 1982), and transferred to Zeta-Probe membrane as described (Chapter 2). After transfer the membrane was

exposed to Fuji X-ray film at room temperature without an intensifying screen.

Since only one end of the phage is labelled, only fragments which are still attached to the right arm are visible on the autoradiograph, and it is possible to deduce where sites for that restriction enzyme are located relative to the labelled right arm.

3.24 Subcloning

Bam HI fragments were subcloned into the *Bam* HI site of the vectors pTZ18U and pTZ19U, as described in the Materials and Methods section of Chapter 2.

3.3 Results

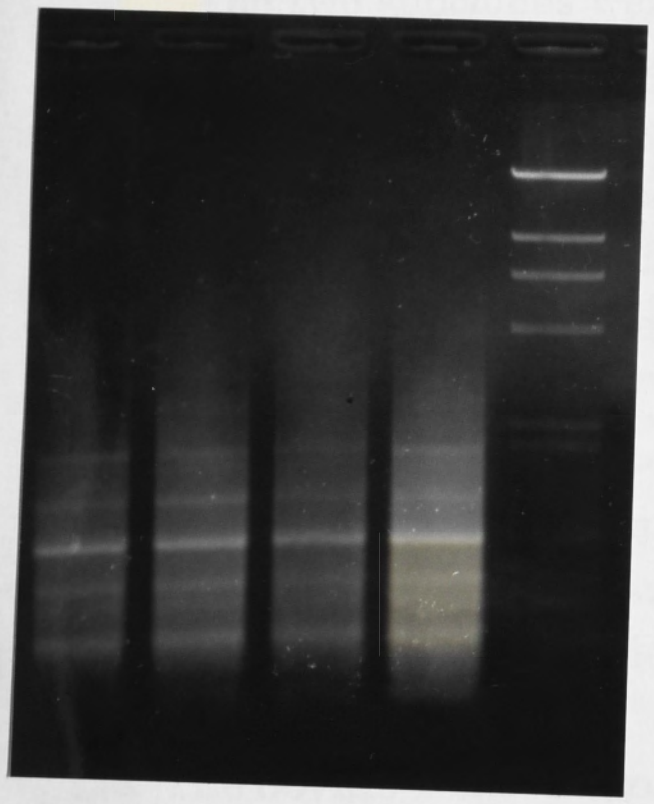
3.31 Isolation of 13 kb from the bovine Y chromosome

Eight putative positives were isolated, as described above from a screen of 240,000 plaques. Phage recombinants containing sequences homologous to BRY.1 were identified by digesting the phage DNA with several restriction enzymes (*Sal* I, *Bam* HI, *Eco* RI and *Hind* III), electrophoresing the fragments in a 0.7% agarose gel with end-labelled λ DNA markers and transferring to a Zeta-Probe membrane. This membrane was probed with nick-translated BRY.1.

Of the eight phage isolated only three identical recombinants hybridized with BRY.1. Two of these were originally from the same isolate, but the recovery of two from a screen of 240,000 shows that the library seems to be reasonably representative.

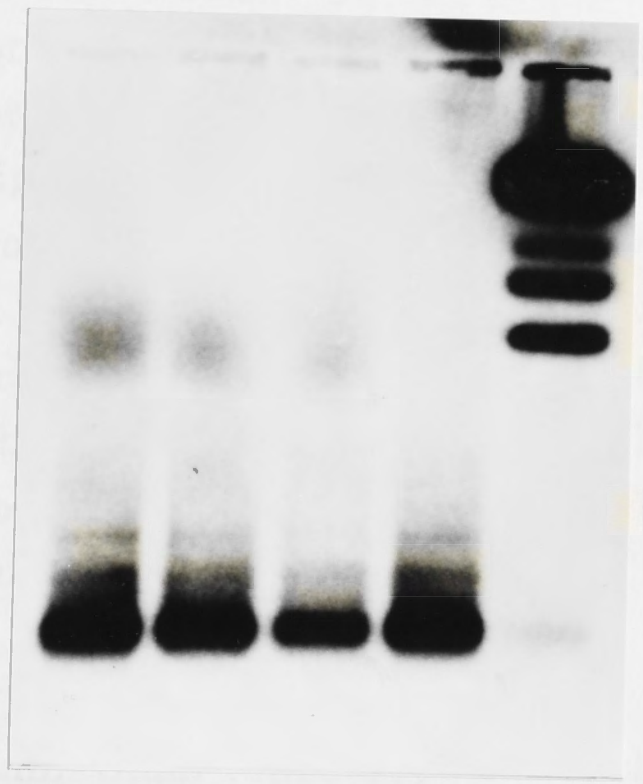
This phage (EMBL3A.Y1) contains an insert of 13 kb of bovine DNA. The phage DNA was labelled by nick-translation and used to probe a Southern blot of *Sau* 3AI-digested male and female cattle DNAs. The bovine insert hybridized to a smear of DNA fragments in both sexes, in a size range of about 300 bp to about 2 kb, with several distinct bands within the smear (Figure 3.1).

Since BRY.1 is known to have one or two homologues on the autosomes and/or X chromosome it was necessary to confirm that the



1 2 3 4 5

3.1A



1 2 3 4 5

3.1B

Figure 3.1. *Sau* 3A1 digested male (1 and 3) and female (2 and 4), cattle DNAs were electrophoresed in a 1% agarose gel, (4 ug/lane), with endlabelled λ *Hind* III markers (lane 5), (**3.1A**), then transferred by alkaline blotting to Zeta-Probe. The filter was hybridized with EMBL3A.Y1 after labelling by nick-translation with [α-32P]-dCTP, (**3.1B**).

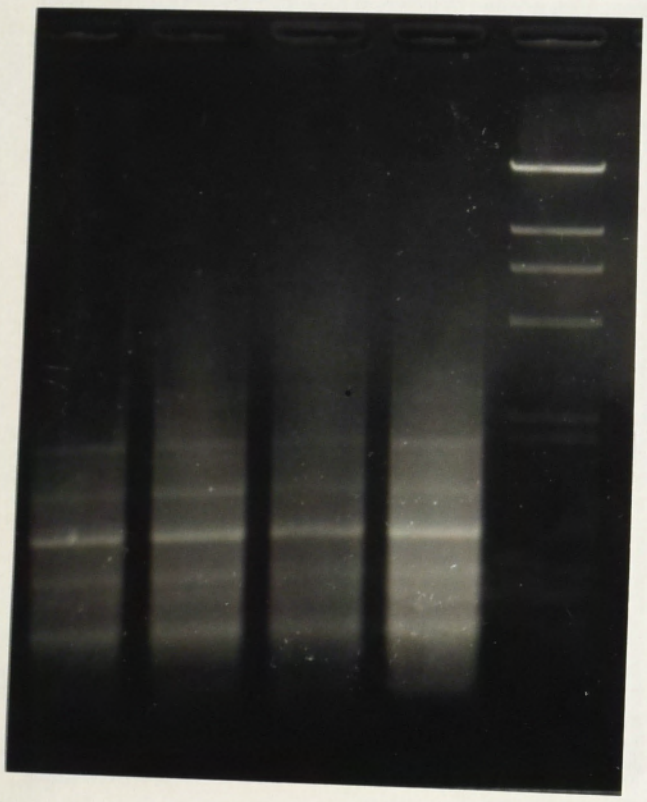
phage contained DNA from the Y chromosome rather than one of the rarer non-Y-chromosomal copies of BRY.1. *Bam* HI digestion of the phage produces four fragments, the two smallest of which hybridize to BRY.1. These fragments (3.7 kb and 4.2 kb) were subcloned into pGEM 1 (Promega) in the hope that they would contain mostly male-associated sequences which would allow a certain identification of Y-chromosomal origin. Both the 3.7 kb and 4.2 kb bands are seen when BRY.1 is used to probe *Bam* HI genomic digests of male cattle DNA (Chapter 2).

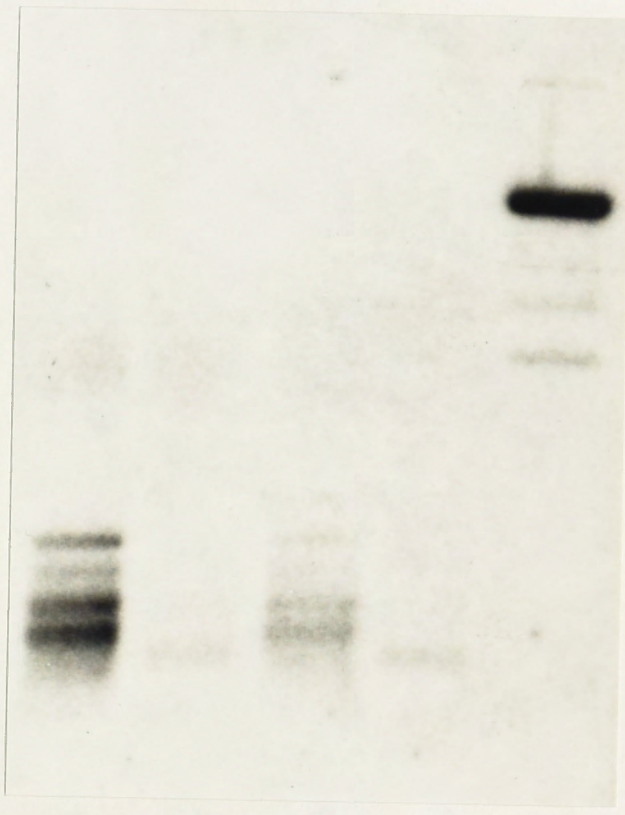
The two subclones were nick-translated and hybridized to Southern blots of *Sau* 3A1 digested cattle genomic DNAs. As can be seen in Figures 3.2 and 3.3 each of the subclones hybridizes preferentially to male *Sau* 3A1 fragments. The smaller subclone (Figure 3.2) hybridizes to four distinct bands in bull DNA, ranging in size from about 300 to about 700 bp,with a female specific band at about 400 bp. This may indicate homology with an X-chromosomal sequence. The other subclone (Figure 3.3) also hybridizes to a range of *Sau* 3A1 fragments in the male, but the most distinct band is probably smaller than 300 bp and the pattern of hybridization is different from that obtained for the smaller subclone. This fragment also shows some hybridization to female DNA.

3.32 The bovine Y chromosome contains other repeats specific to the Y chromosome

The pattern of hybridization of both subclones is different to that of BRY.1, which hybridizes strongly to a single 300 bp *Sau* 3AI fragment and weakly to several larger fragments in male DNA (Figure 2.1, Chapter 2), suggesting each subclone contains additional repeats specific to the Y chromosome and also additional non-Y-specific repeats. The pattern of hybridization is also different for each subclone which implies that each contains different Y chromosomal repeats as well as BRY.1.

Having confirmed that the phage does contain DNA from the cattle Y chromosome the smaller (3.7 kb) fragment was designated BRY.2, the larger (4.2 kb) fragment, BRY.3. A similarly male-associated result was obtained (Figures 3.4 and 3.5) when the subclones were labelled by nick-translation and hybridized to Southern blots of *Bam* HI digested male and female cattle DNA. A striking difference was seen between the hybridization pattern produced by each subclone and BRY.1 and between





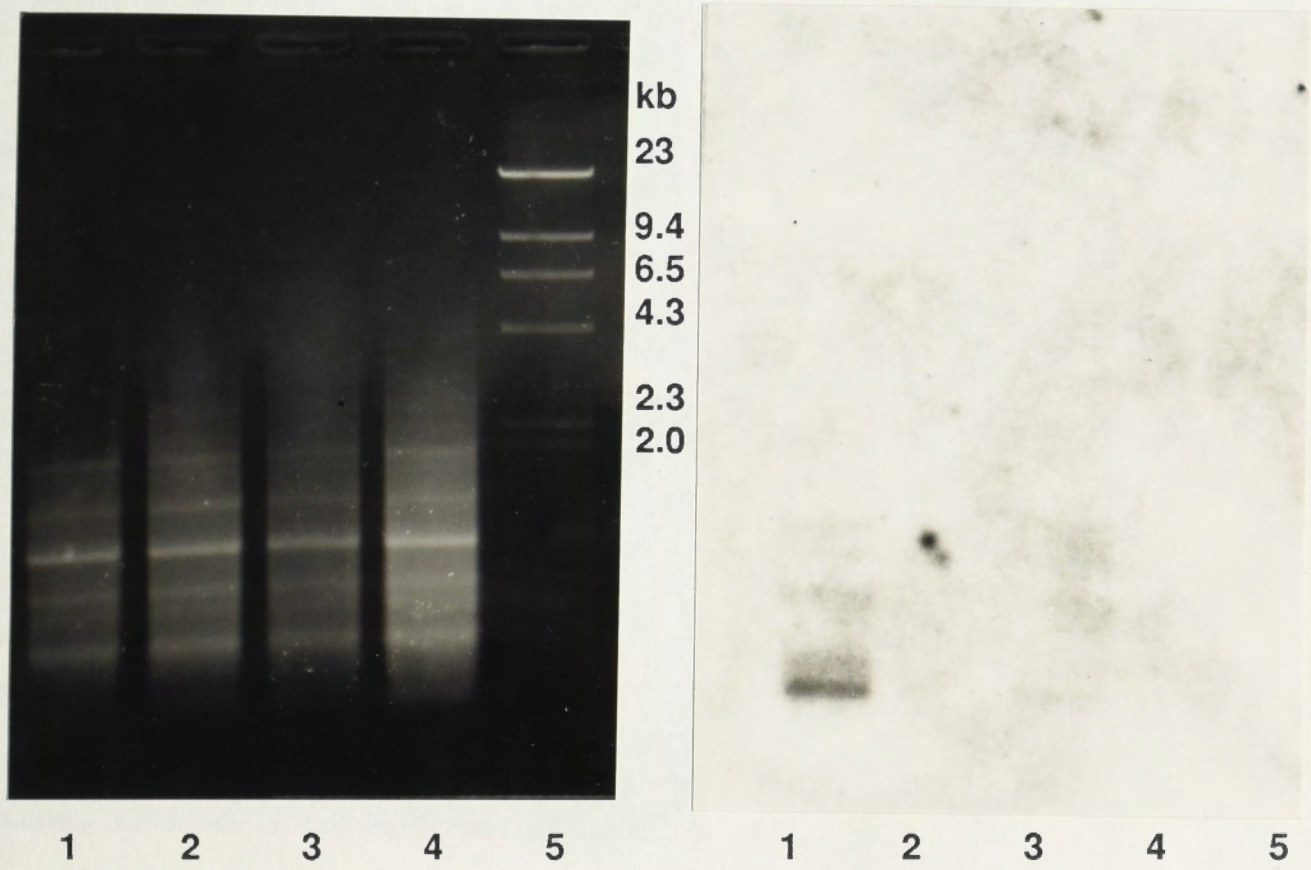
1 2 3 4 5

3.2A

1 2 3 4 5

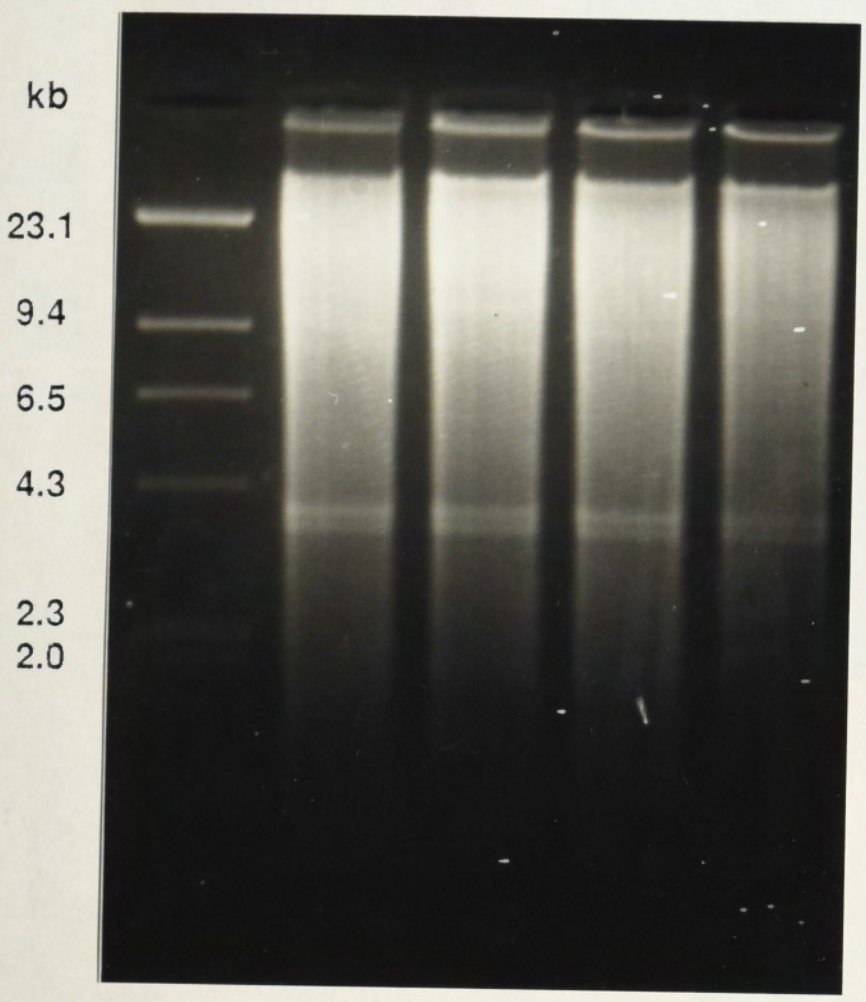
3.2B

Figure 3.2. *Sau* 3A1digested male (lanes 1 and 3) and female (2 and 4), cattle DNAs were electrophoresed in a 1% agarose gel (4 ug/lane), with endlabelled λ *Hind* III markers (lane 5), (**3.2A**), then transferred by alkaline blotting to Zeta-Probe. The filter was probed with pG.BRY.2, labelled by nick-translation (**3.2B**).

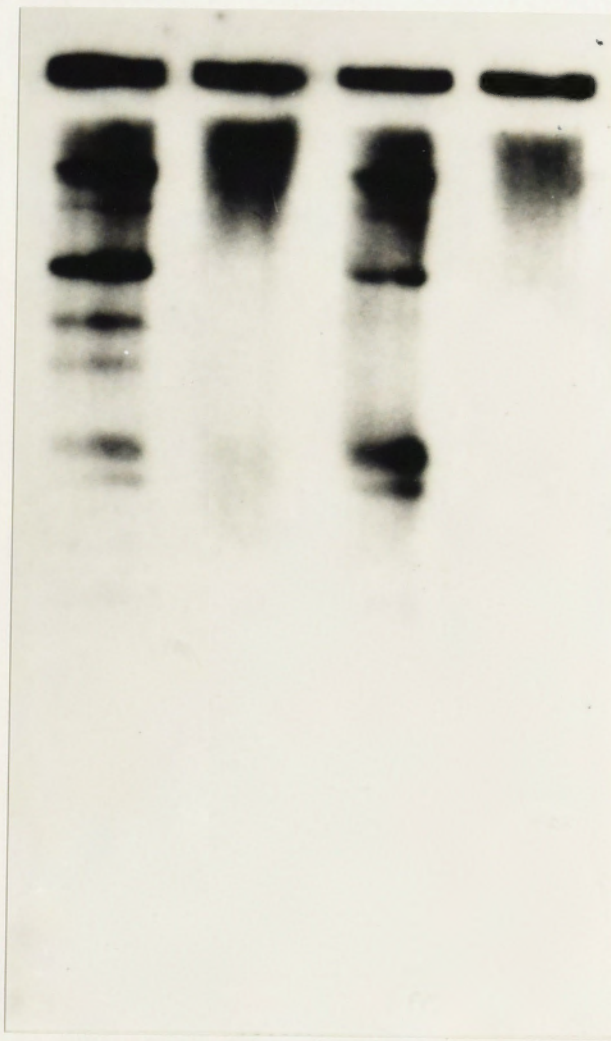


3.3A 3.3B

Figure 3.3. *Sau* 3A1 digested male (lanes 1 and 3) and female (2 and 4), cattle DNAs were electrophoresed in a 1% agarose gel (4 ug/lane), with endlabelled λ *Hind* III markers (lane 5), (**3.3A**), then transferred by alkaline blotting to Zeta-Probe. The filter was probed with pG.BRY.3, labelled by nick-translation (**3.3B**).



3.4A



3.4B

Figure 3.4. *Bam* HI digested male (lanes 2 and 4) and female (3 and 5), cattle DNAs were electrophoresed in a 0.6% agarose gel, (4 µg/lane), with λ *Hind* III markers (lane 1), (**3.4A**), then transferred by alkaline blotting to Zeta-Probe. The filter was hybridized with pG.BRY.2, labelled by nick-translation (**3.4B**).

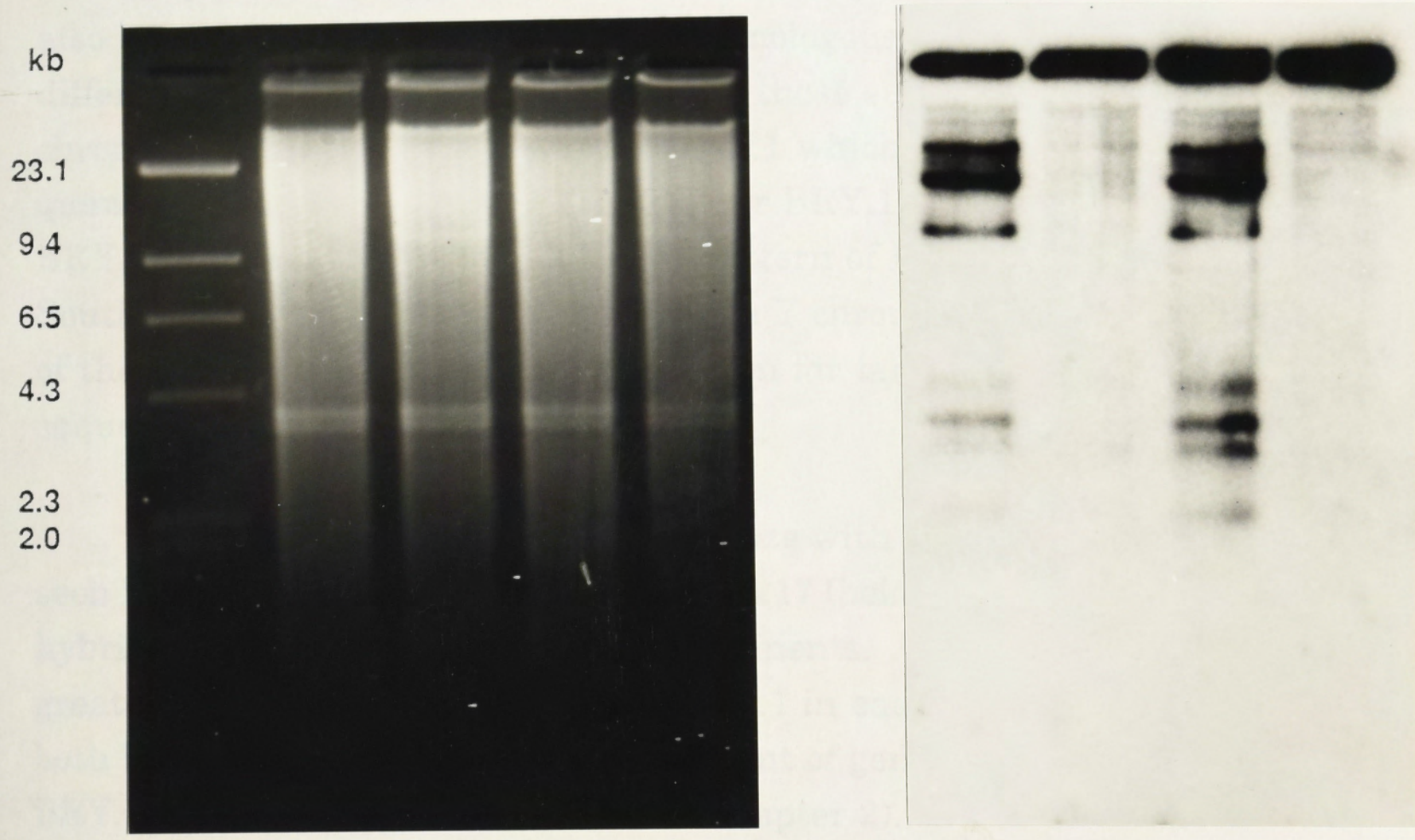


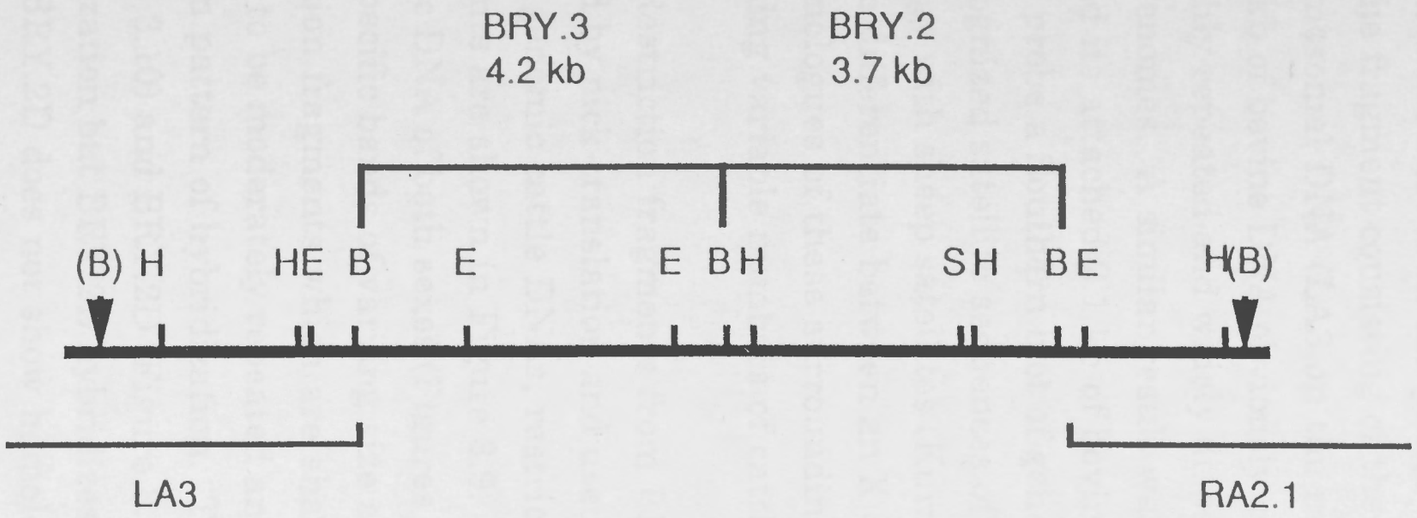
Figure 3.5. *Bam* HI digested male (lanes 2 and 4) and female (3 and 5), cattle DNAs were electrophoresed in a 0.6% agarose gel, (4 µg/lane), (**3.5A**), then transferred by alkaline blotting to Zeta-Probe. The filter was hybridized with pG.BRY.3, labelled by nick-translation (**3.5B**). Lane 1 contains λ *Hind* III markers.

the two subclones. BRY.2 hybridizes to *Bam* HI fragments of 19, 10, 6.7, 6.5, 4.2, 3.7, 3.5, 2.2, 2, and 1.8 kb in size in male DNA (Figure 3.4), while BRY.3 hybridizes to fragments of 19, 9.5, 4.8, 4.2, 3.7, 3.2, 1.7, and 1.6 kb in male DNA (Figure 3.5). While the *Bam* HI fragments of genomic DNA to which BRY.1 hybridizes are bound by each of BRY.2 and BRY.3, each subclone also hybridizes to several bands non-homologous to BRY.1 which are different for BRY.2 and BRY.3. Each of these subclones contains Y-chromosomal sequences other than BRY.1 which appear to be repeated and quantitatively polymorphic. As found for BRY.1, sequences homologous to BRY.2 and BRY.3 appear, from their pattern of hybridization to genomic Southern blots, to be dispersed along the Y chromosome, with no evidence of the regular 'ladder' effect which is seen for tandemly repeated sequences.

The two subclones do cross-hybridize with one another, as can be seen in the dot blots of Figures 3.16 and 3.17 (below), as expected from their hybridization to common restriction fragments. The homology seems to be greater than simply the presence of BRY.1 in each of the subclones, since both hybridize to a 19 kb *Bam* HI fragment of genomic bull DNA, to which BRY.1 does not hybridize (Figure 2.3, Chapter 2). It can also be seen from figures 3.2, 3.3, 3.4 and 3.5 that both BRY.2 and BRY.3 contain sequences which are found elsewhere in the genome.

3.33 Y-specific sequences are interspersed with repeats found elsewhere in the genome

The position of the *Bam* HI fragments, subcloned to give BRY.2 and BRY.3, within the original phage EMBL3A.Y1 were determined using Klenow labelling of the right *cos* arm as described in Materials and Methods. Figure 3.6 shows the ordering of the *Bam* HI fragments and several other restriction sites mapped as a preliminary to sequencing. From this restriction map it can be seen that BRY.2 and BRY.3, containing mostly repeated sequences specific to the Y chromosome (including BRY.1), are in the centre of the 13 kb of cloned Y chromosomal DNA. The pattern of hybridization seen when the whole phage is used as a probe (Fig. 3.1) shows no sex discrimination at all, which coupled with the data from Figures 3.2 and 3.3, suggests that one or both of the pieces of Y chromosomal DNA either side of this Y-specific block contains sequences which are found in



Scale: 1 cm=1kb

(B)- *Bam* HI site not regenerated
B - *Bam* HI site
H - *Hind* III site
E - *Eco* RI site
S - *Sac* I site

Figure 3.6 . Restriction map of EMBL3A.Y1 isolated using BRY.1, showing the subclones BRY.2 and BRY.3 and the regions used to probe Figures 3.7 and 3.8.

moderate to high copy number, and which are dispersed on the autosomes and/or the X chromosome, as well as the Y chromosome.

Each of the larger *Bam* HI fragments from EMBL3A.Y1, representing the two arms of the phage with attached bovine DNA, was isolated by gel purification, labelled by nick-translation, and used to probe a Southern blot of *Bam* HI digested genomic cattle and sheep DNAs. These fragments are shown on the phage map (Figure 3.6). The result obtained using the fragment consisting of the left arm of the phage and 3 kb of Y-chromosomal DNA (LA3 on the map, Figure 3.6) is shown in Figure 3.7. This 3 kb of bovine DNA obviously contains a sequence or sequences which are highly repeated and widely dispersed throughout both the bovine and ovine genomes. A similar result was obtained (Figure 3.8) when the right arm and its attached 2.1 kb of bovine Y-chromosomal DNA, RA2.1, was used to probe a Southern blot of genomic DNAs. This sequence is not one of the recognized satellite sequences of the cattle genome since they have no homology with sheep satellites (Kurnit *et al.*, 1978). It is impossible at this stage to differentiate between an X chromosomal or autosomal location for the homologues of these surrounding sequences because cell lines containing variable numbers of cattle X chromosomes are not available.

Restriction fragments from BRY.2 were subcloned into pGEM1, labelled by nick-translation and used to probe Southern blots of male and female genomic cattle DNAs, restricted with *Hind* III. These restriction fragments are shown in Figure 3.9. Three of these fragments hybridized to genomic DNA of both sexes (Figures 3.10, 3.11 and 3.12), but with distinct male-specific bands of varying size and intensity. The sequences in these restriction fragments which are shared by the male and female genomes appear to be moderately repeated and related to one another, having a common pattern of hybridization. The restriction fragments BRY.2B (Figure 3.10) and BRY.2D (Figure 3.12) show a common pattern of hybridization but BRY.2B hybridizes to an extra male-specific band to which BRY.2D does not show homology. BRY.1 is found within the restriction fragment BRY.2D. One of these subcloned BRY.2 fragments (designated BRY.2A in Figure 3.9) produced a very male-specific pattern as seen in Figure 3.13. This restriction fragment of 700 bp does not contain BRY.1,as found by hybridization analysis (and later confirmed by sequencing, Chapter 5). Therefore BRY.2 contains at least two, and

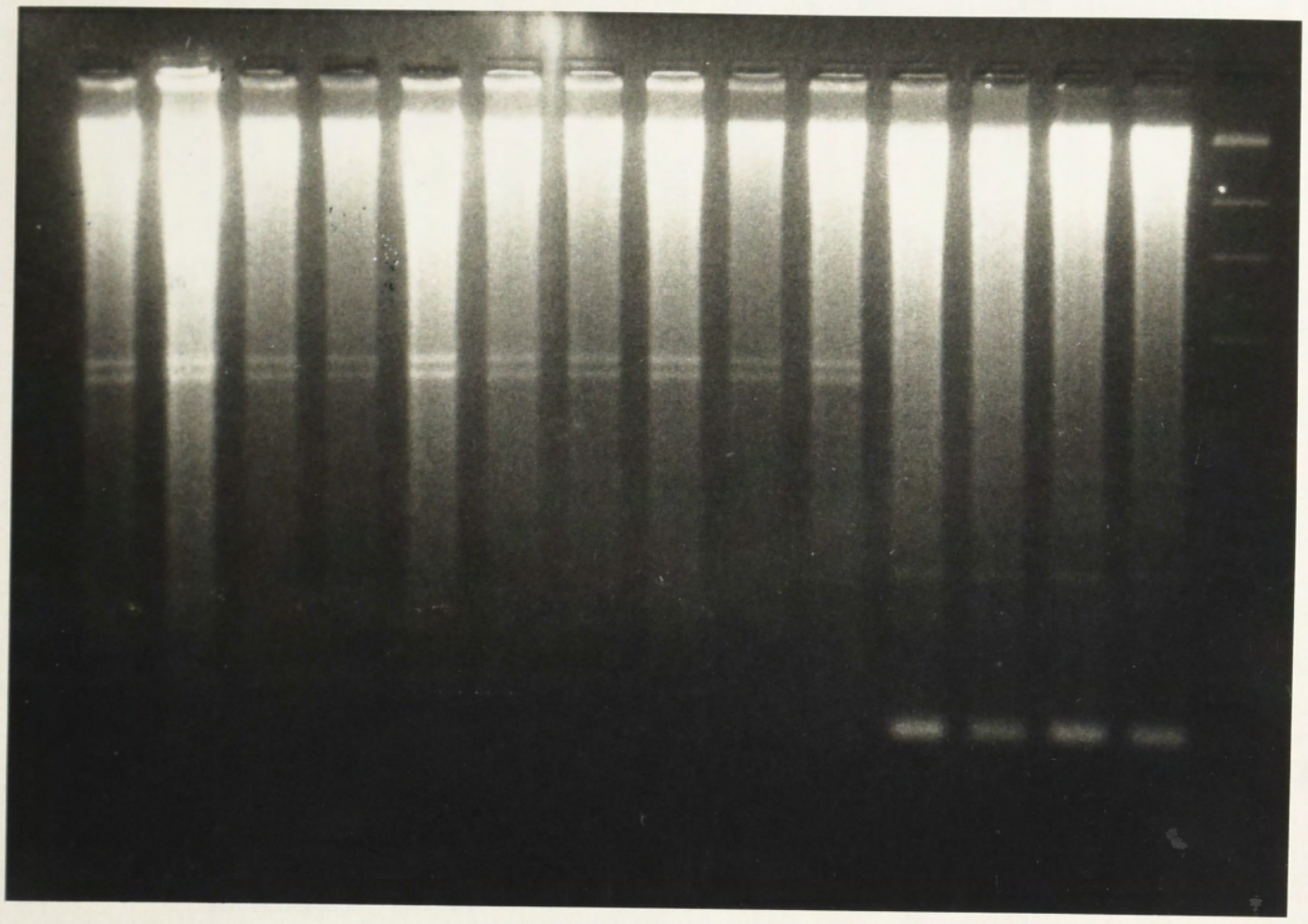
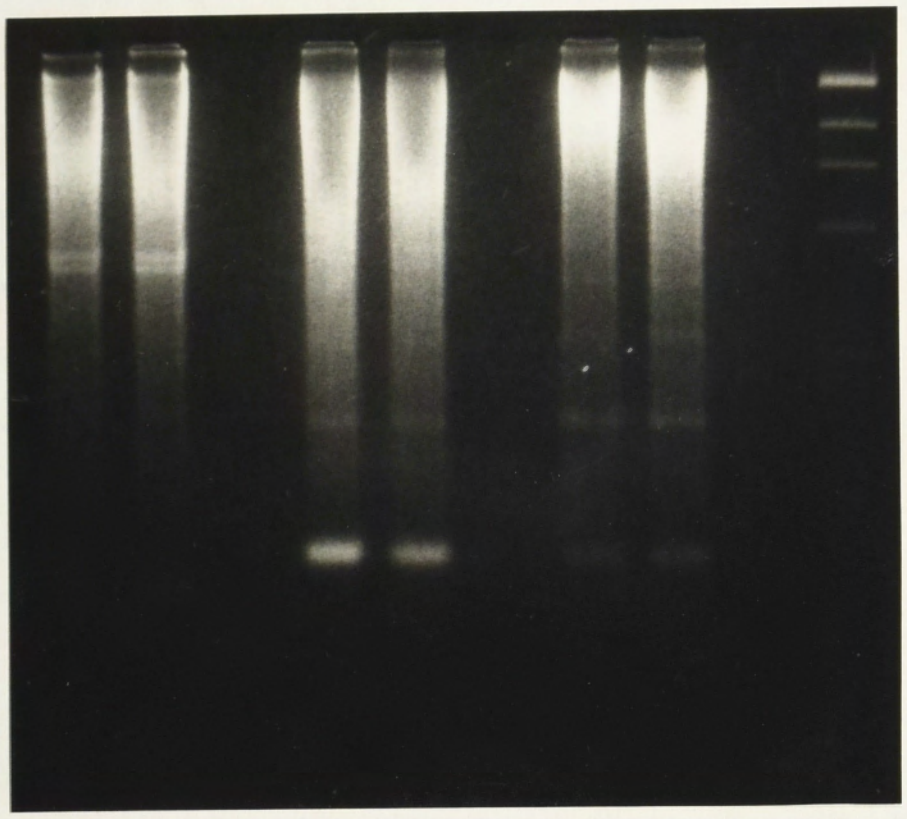


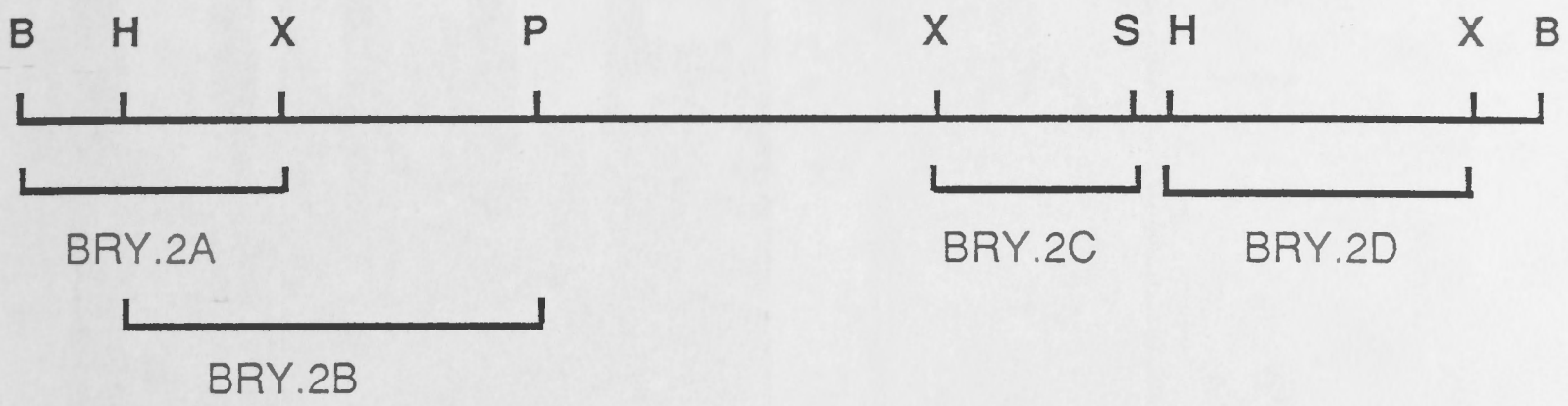
Figure 3.7. *Bam* HI digested genomic cattle male (lanes 1,4, 9 and 10), and female (lanes 2,3,5,6,7,8,) and sheep male (lanes 11 and 13), and female (lanes 12 and 14), were electrophoresed in a 1% agarose gel (**3.7A**), with λ *Hind* III markers (lane 15), then transferred to Zeta-Probe. The membrane was hybridized with the *Bam* HI fragment LA3 (see Fig. 3.6), which was labelled by nick-translation. The membrane was exposed to Fuji-RX overnight at room temperature. The autoradiograph (**3.7B**), shows this fragment contains a sequence which is present at high copy numbers in both cattle and sheep genomes and which is highly dispersed throughout these genomes.



3.8A

3.8B

Figure 3.8. *Bam* HI digested genomic cattle male (lane1)and female (lane 2), sheep male (lane 3) and female (lane 4), and goat male (lane 5) and female (lane 6) DNAs were electrophoresed in a 1% agarose gel (**3.8A**), with endlabelled λ *Hind* III markers (lane 7), then transferred to Zeta-Probe. The membrane was hybridized with the *Bam* HI fragment RA2.1 (see Fig. 3.6), which was labelled by nick-translation. The autoradiograph (**3.8B**), shows this fragment contains a sequence which, is present at high copy numbers in both cattle and sheep genomes and which is highly dispersed throughout these genomes, giving the same result as for LA3 (Fig. 3.7A). The sequence is however present at a much lower level in the goat genome. The membrane was exposed to Fuji-RX for four hours at room temperature.



B -*Bam* HI site
H - *Hind* III site
X - *Xba* I site
S -*Sac* I site
P - *Pst* I site

4 cm = 1 kb

Figure 3.9. Restriction fragments from BRY.2 used to probe Southern blots of genomic DNAs in Figures 3.10, 3.11, 3.12 and 3.13.

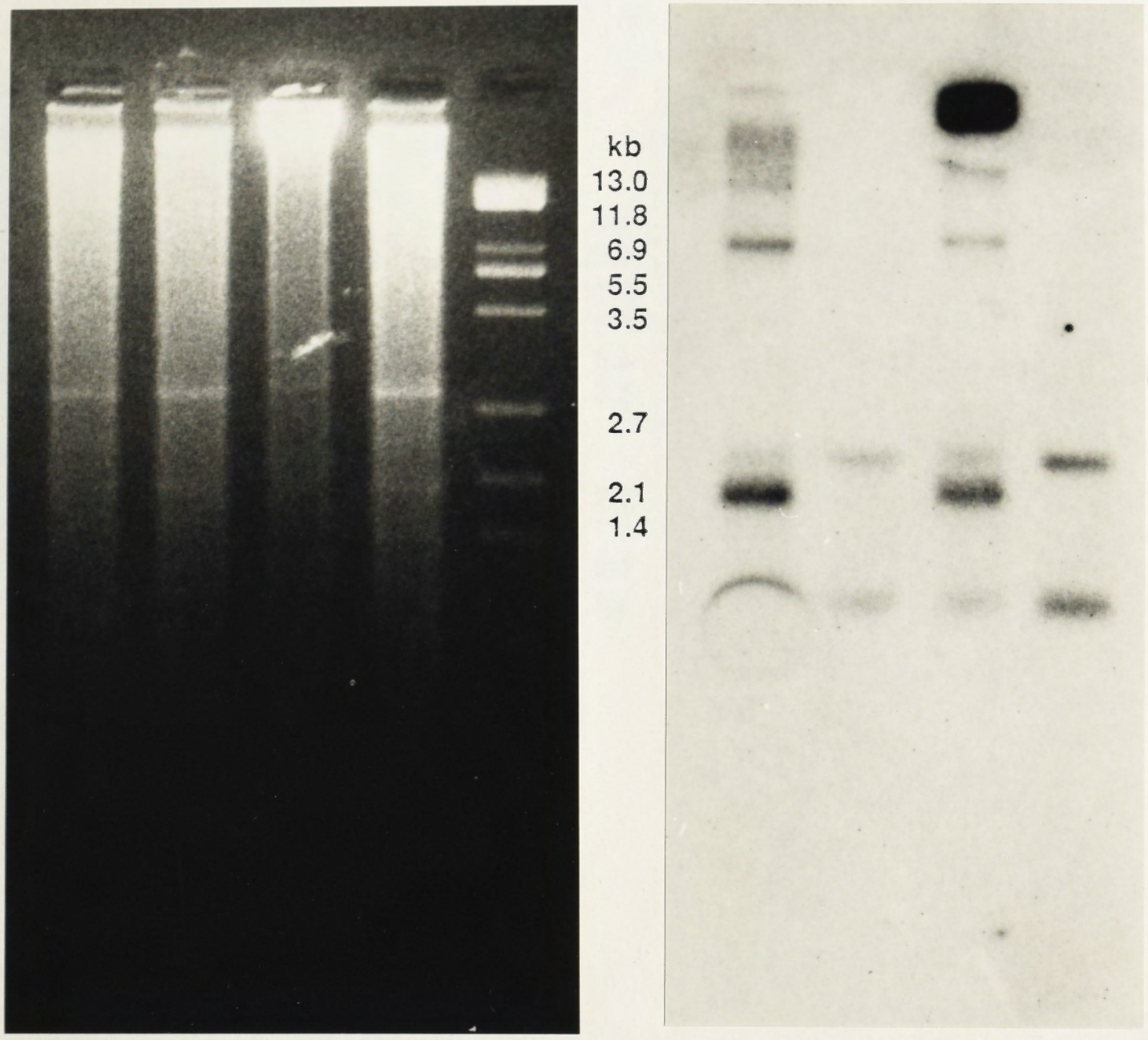
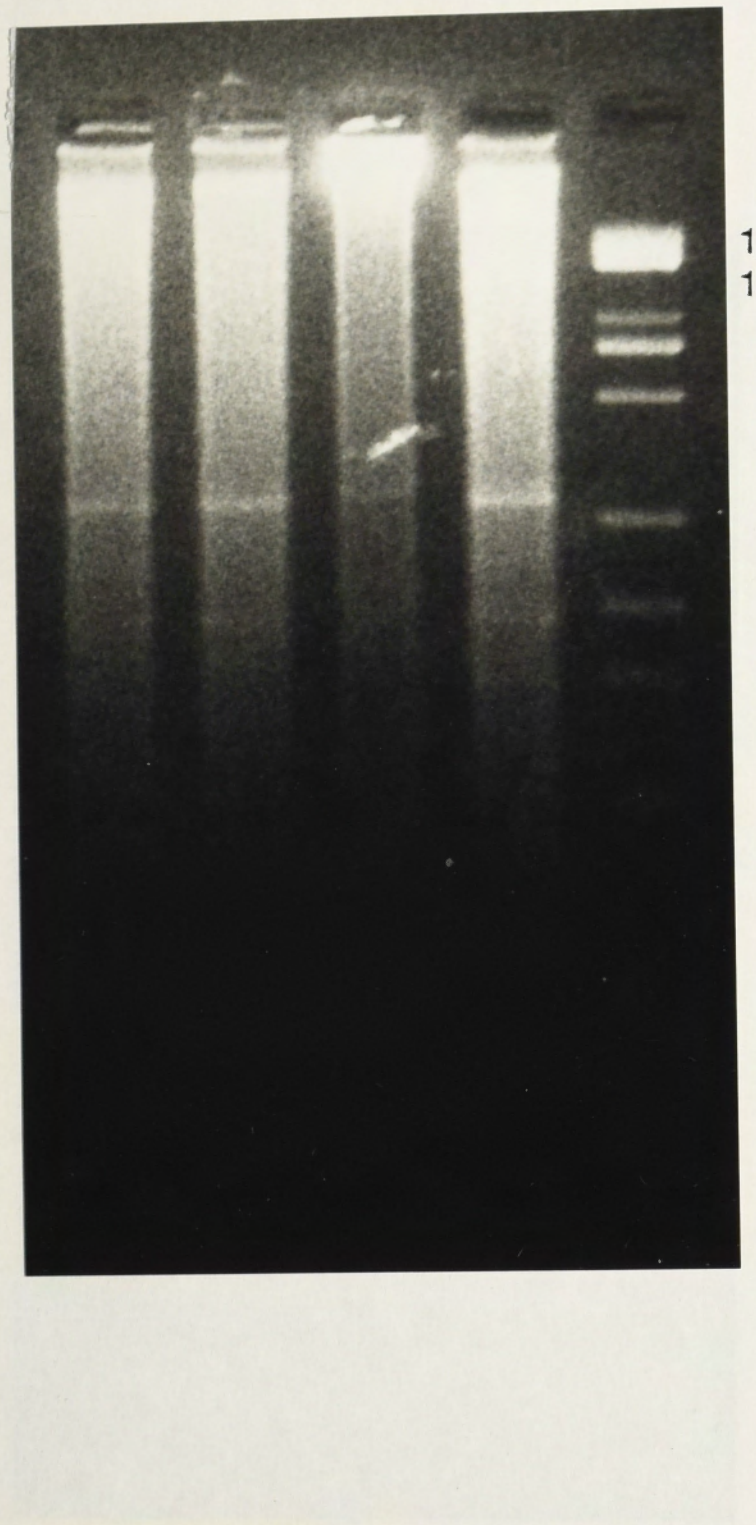


Figure 3.10. *Hind* III digested male (lanes 1 and 3) and female (2 and 4) cattle DNAs were electrophoresed in a 0.8% agarose gel (**3.10A**), then transferred by alkaline blotting to Zeta-Probe. Lane 5 contains λ *Acc* I markers. The filters were probed with nick-translated BRY.2B (**3.10B**). The position of BRY.2B wiithin BRY.2 can be seen in Figure 3.9.



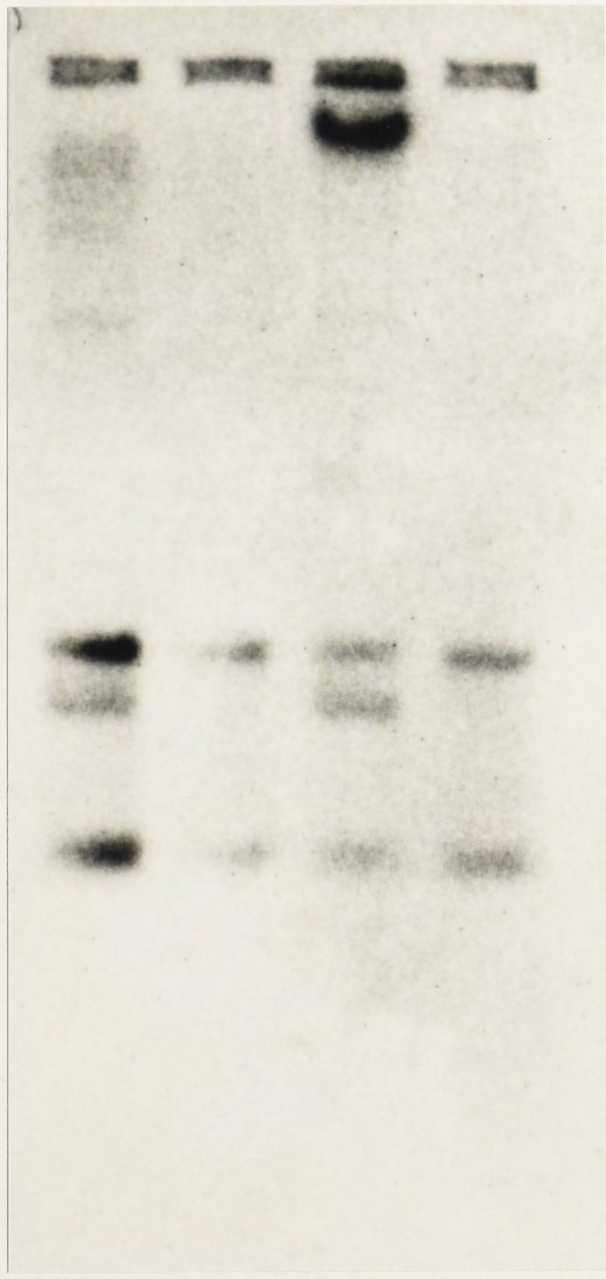
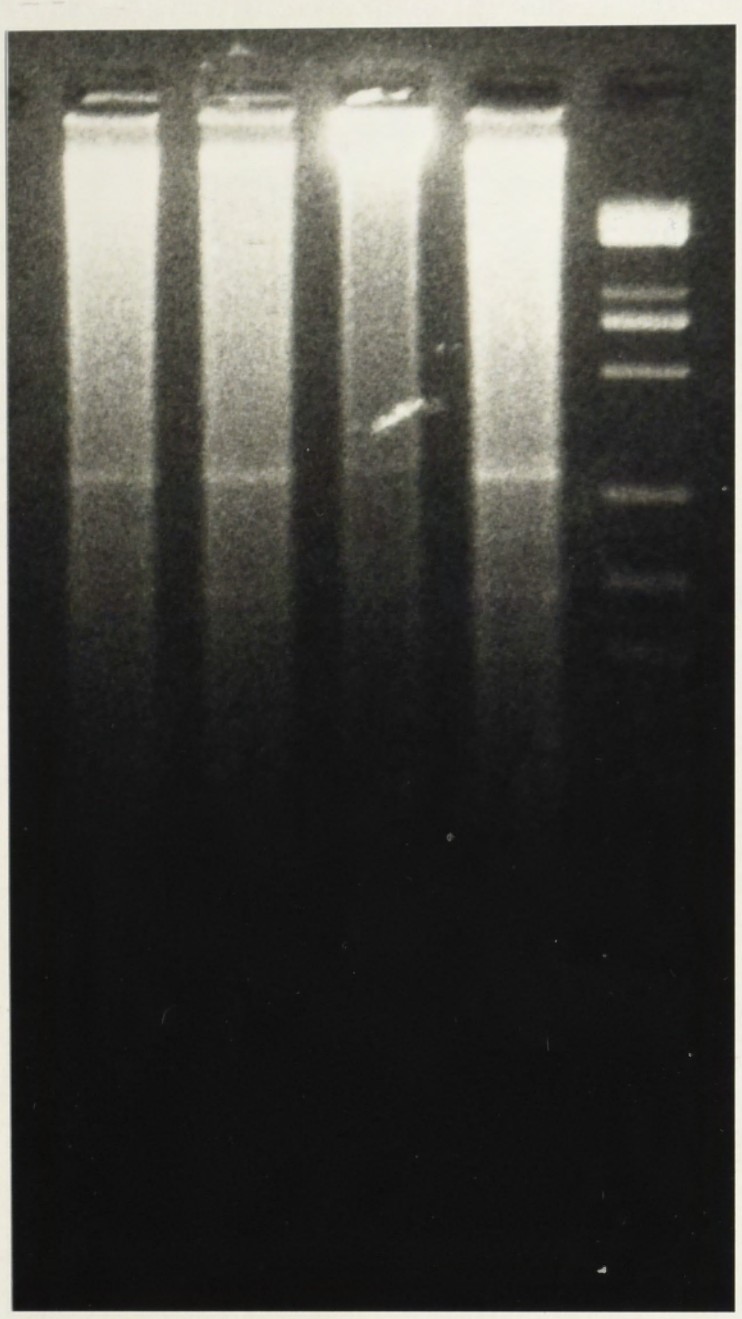
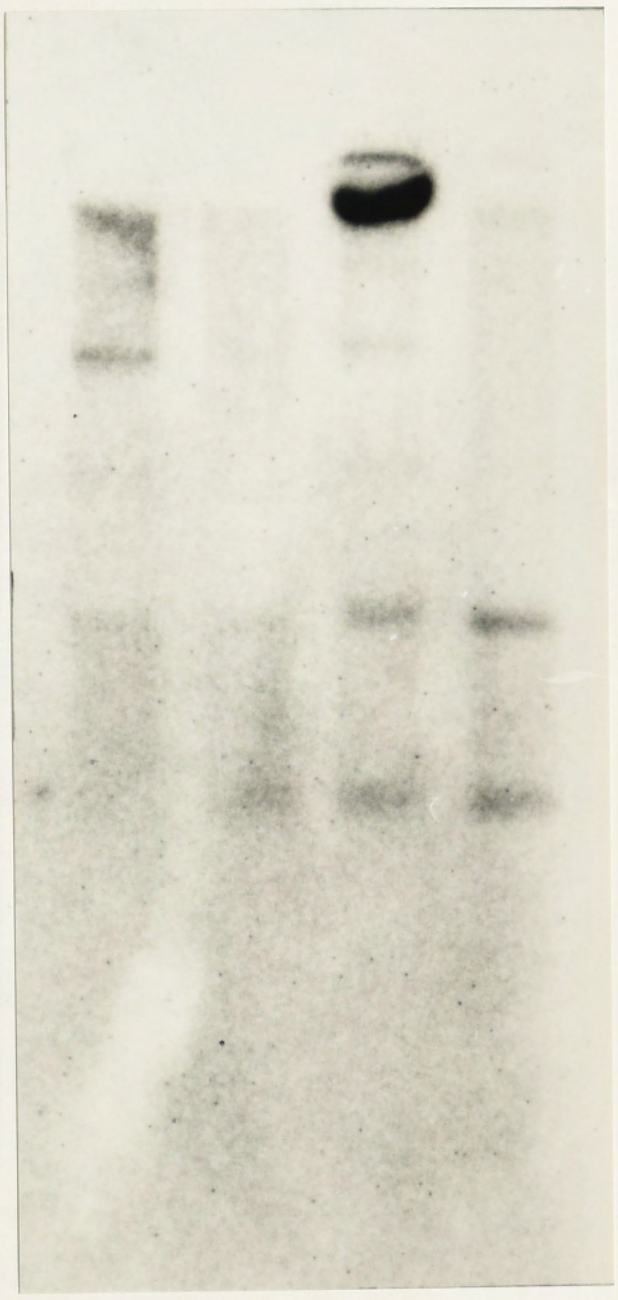


Figure 3.11. *Hind* III digested male (lanes 1 and 3) and female (2 and 4) cattle DNAs were electrophoresed in a 0.8% agarose gel (**3.11A**), then transferred by alkaline blotting to Zeta-Probe. Lane 5 contains λ *Acc* I markers. The filters were probed with nick-translated BRY.2C (**3.11B**). The position of BRY.2C wiithin BRY.2 can be seen in Figure 3.9.



1 2 3 4 5

3.12A



1 2 3 4

3.12B

Figure 3.12. *Hind* III digested male (lanes 1 and 3) and female (2 and 4) cattle DNAs were electrophoresed in a 0.8% agarose gel (**3.12A**), then transferred by alkaline blotting to Zeta-Probe. Lane 5 contains λ *Acc* I markers. The filters were probed with nick-translated BRY.2D (**3.12B**). The position of BRY.2D wiithin BRY.2 can be seen in Figure 3.9.

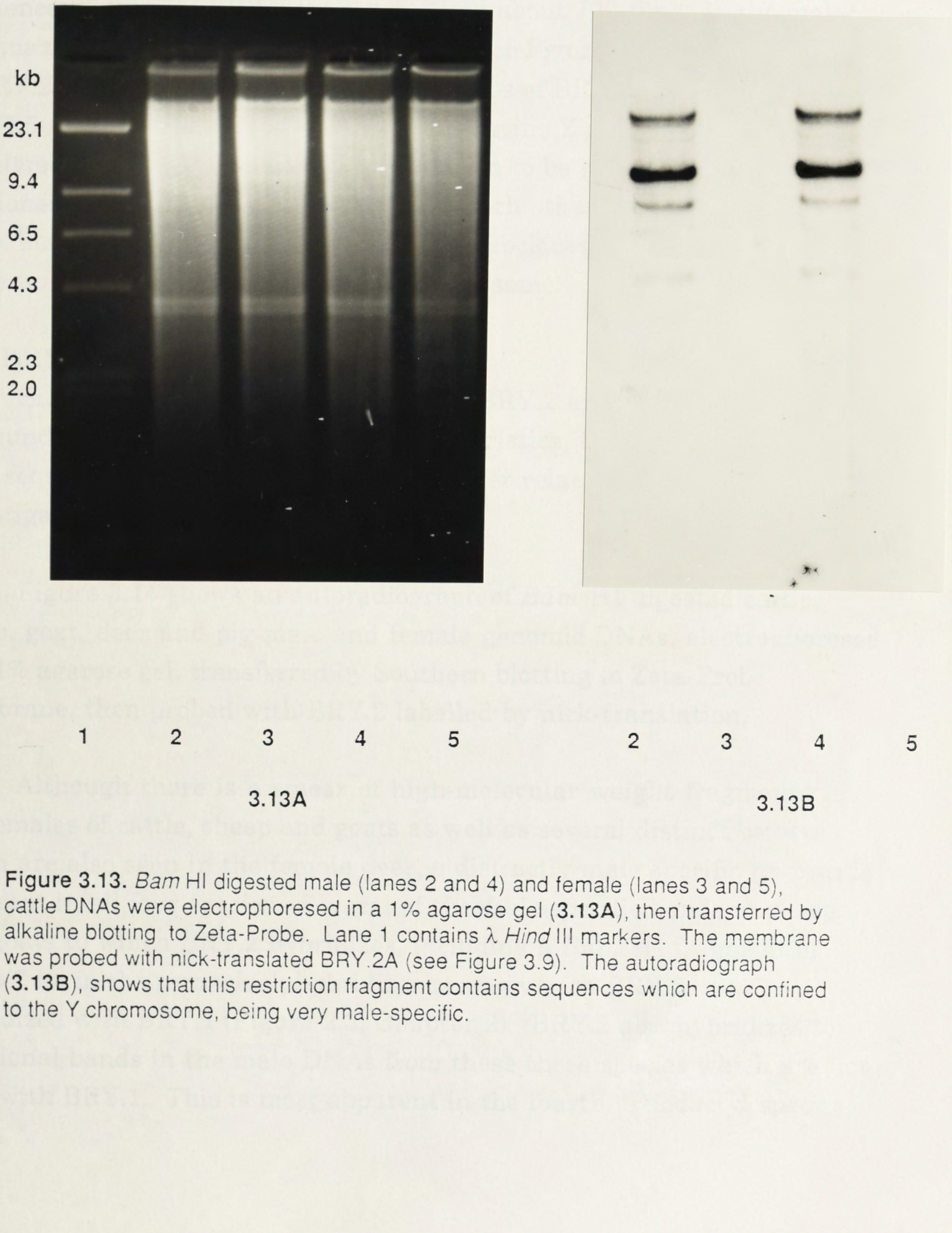


Figure 3.13. *Bam* HI digested male (lanes 2 and 4) and female (lanes 3 and 5), cattle DNAs were electrophoresed in a 1% agarose gel (**3.13A**), then transferred by alkaline blotting to Zeta-Probe. Lane 1 contains λ *Hind* III markers. The membrane was probed with nick-translated BRY.2A (see Figure 3.9). The autoradiograph (**3.13B**), shows that this restriction fragment contains sequences which are confined to the Y chromosome, being very male-specific.

possibly three, different male-specific repeated sequences. The male-specific nature of some of the sequences in the other subcloned restriction fragments may be obscured by the inclusion of a neighbouring non-Y-specific sequence in the subclone.

3.34 Sequences related to BRY.2 and BRY.3 comprise 40% of the cattle Y

The copy number of sequences related to some part of BRY.2 was estimated by quantitative dot blot (see Figure 3.16) to be approximately 1200 copies per genome in the male, and about 100 copies in the female. Sequences similar to BRY.3 are represented about 700 times in the male genome and about 100 times in the female (see Figure 3.17). The 1100 copies of BRY.2 (or parts thereof) with the 600 copies of BRY.3 account for $100/(1.5 \times 10^7 \text{ bp}/ 6.6 \times 10^6 \text{ bp}) = 44\%$ of the bovine Y chromosome. Some of the elements in BRY.2 and BRY.3 are known to be shared since both subclones hybridize with BRY.1 and with each other (Figures 3.16 and 3.17), but the elements found in these two subclones from a single phage make up at least 40% of the cattle Y chromosome.

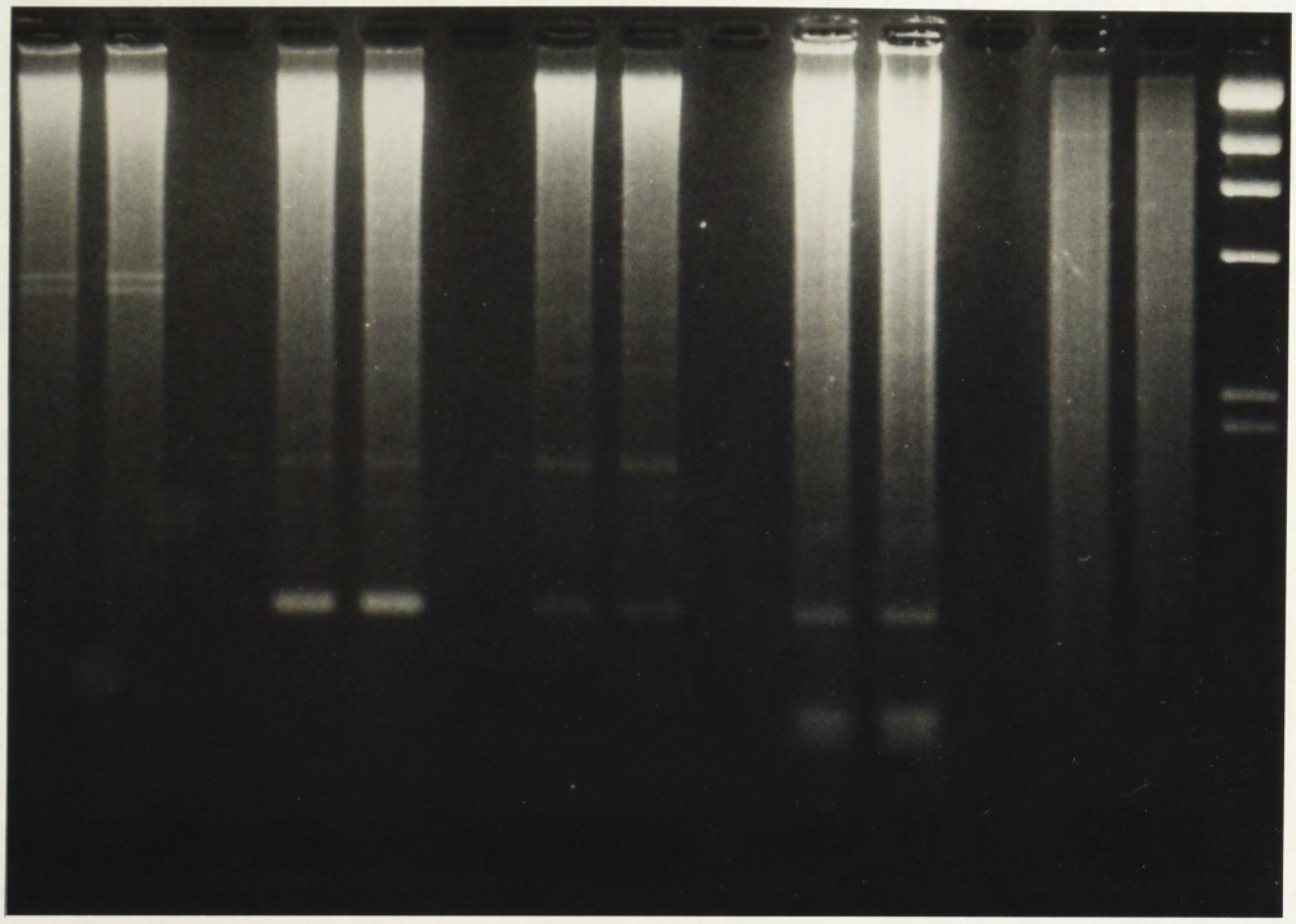
3.35 BRY.2 and BRY.3 in other species

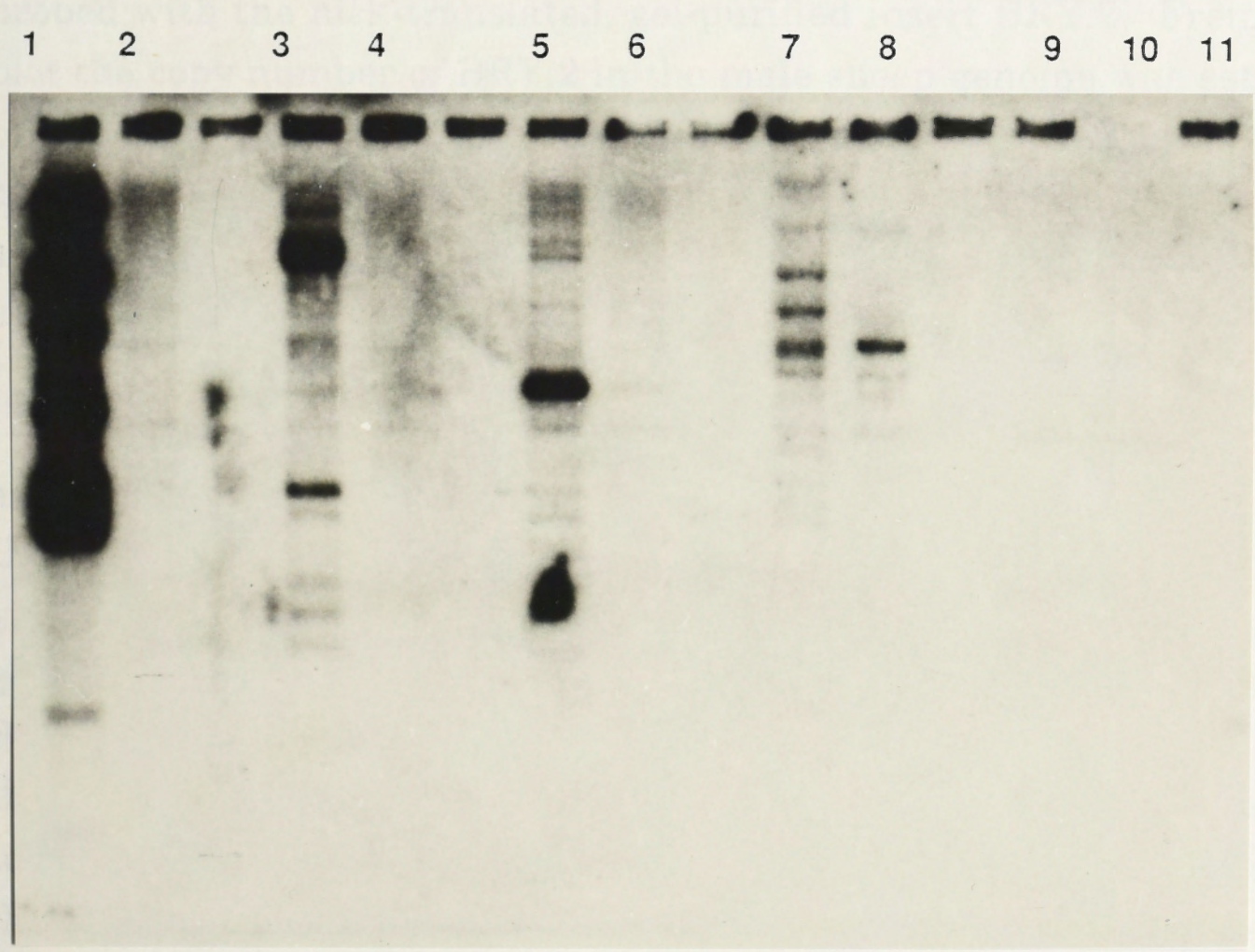
Since the Y-associated sequences in BRY.2 and BRY.3 which surround BRY.1 share many of its characteristics, the extent of homology with sequences on the Y chromosome of other related species was investigated.

Figure 3.14 shows an autoradiograph of *Bam* HI digested cattle, sheep, goat, deer and pig male and female genomic DNAs, electrophoresed in a 1% agarose gel, transferred by Southern blotting to Zeta-Probe membrane, then probed with BRY.2 labelled by nick-translation.

Although there is a smear of high-molecular weight fragments in the females of cattle, sheep and goats as well as several distinct bands which are also seen in the female deer, a distinctly male specific pattern is apparent in all four species. Several of the male-specific bands in sheep and goats to which BRY.2 hybridizes, including a *Bam* HI fragment of about 7 kb in the ram which hybridizes strongly with BRY.2, also hybridized with BRY.1 (Figure 2.3, Chapter 2). BRY.2 also hybridizes to additional bands in the male DNAs from these three species which are not seen with BRY.1. This is most apparent in the fourth artiodactyl species,

3.14A





3.14B

Figure 3.14. 2 µg of cattle male and female DNA (lanes 1 and 2), sheep male and female (lanes 3 and 4), goat male and female (lanes 5 and 6), deer male and female (lanes 7 and 8) and pig male and female (lanes 9 and 10) DNAs were digested with *Bam* HI and electrophoresed overnight at 12 volts in a 1% agarose gel, (**3.14A**). The DNAs were transferred by alkaline blotting to Zeta-Probe membrane and probed with nick-translated BRY.2, (**3.14B**). Lane 11 contains λ *Hind* III markers.

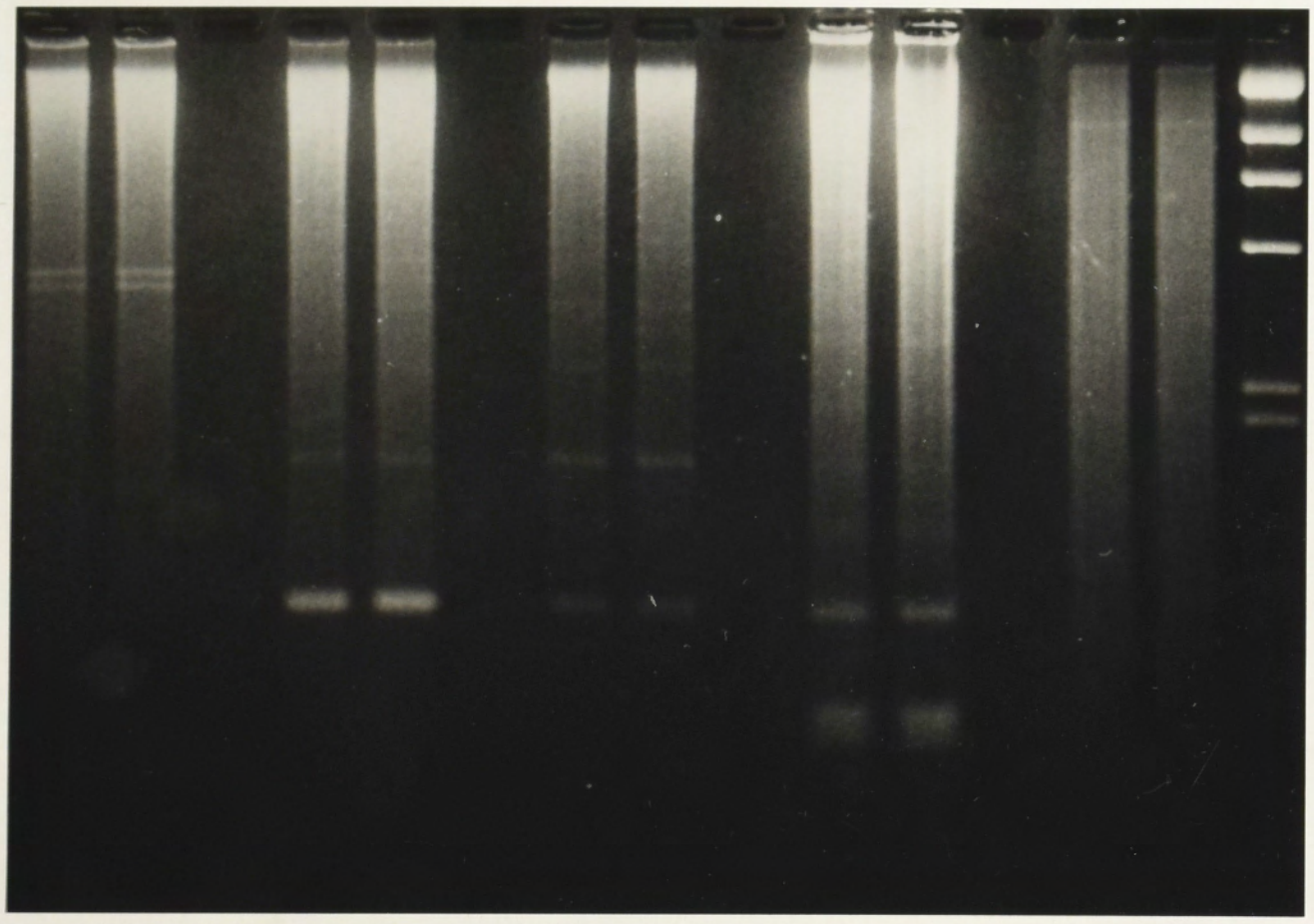
deer. BRY.1 does not discriminate between male and female deer, but BRY.2 contains additional repeated sequences which are male-specific in deer. As for BRY.1 however, BRY.2 hybridizes to a *Bam* HI fragment of about 3 kb in both sexes of pig.

Figure 3.15 shows an identical Southern blot of digested DNAs from male and female cattle, sheep, goats deer and pigs, probed with nick-translated BRY.3. Comparison of Figures 3.14 and 3.15 shows a distinct difference in the patterns of hybridization of the two subclones. BRY.3 also gives a predominantly male-specific pattern in cattle and sheep but no male specificity is seen for goats, deer or pigs. BRY.3 hybridizes more strongly and to a greater number of bands in both sexes of all the species examined than does BRY.2, the Southerns both being exposed to film for three days.

Figure 3.16 is an autoradiograph of a dot blot of cattle, sheep, goat and deer DNA from each sex, with a known amount of the *Sau* 3A1 fragment BRY.1, and the *Bam*HI restriction fragments BRY.2 and BRY.3, probed with the nick-translated, gel-purified insert BRY.2. From this dot blot the copy number of BRY.2 in the male sheep genome was estimated to be approximately 200, with 85 in the female. The male goat genome also contains 200 copies, with the female again having only 85 related sequences. In the male deer there are about 75 copies, while the female has approximately half this number, 35-40.

Copy numbers of BRY.3 in the genome of each species were estimated from the dot blot shown in Figure 3.17. This dot blot is identical to that in Figure 3.17, but has been probed with the nick-translated, gel-purified 4.2 kb *Bam* HI fragment BRY.3. The male sheep and goat genomes were estimated to contain 350 copies of sequences similar to BRY.3, while the female sheep and goat genomes have about 175. The fact that the male goat genome contains twice the number of sequences similar to BRY.3 than the female genome is not apparent from the Southern blot (Figure 3.15). This may indicate that the additional copies on the Y chromosome are within the very large (> 23 kb) *Bam* HI fragments at the top of the gel, which are not resolved. Both sexes of deer have about 350 copies of these sequences.

3.15A



1 2 3 4 5 6 7 8 9 10 11

3.15B

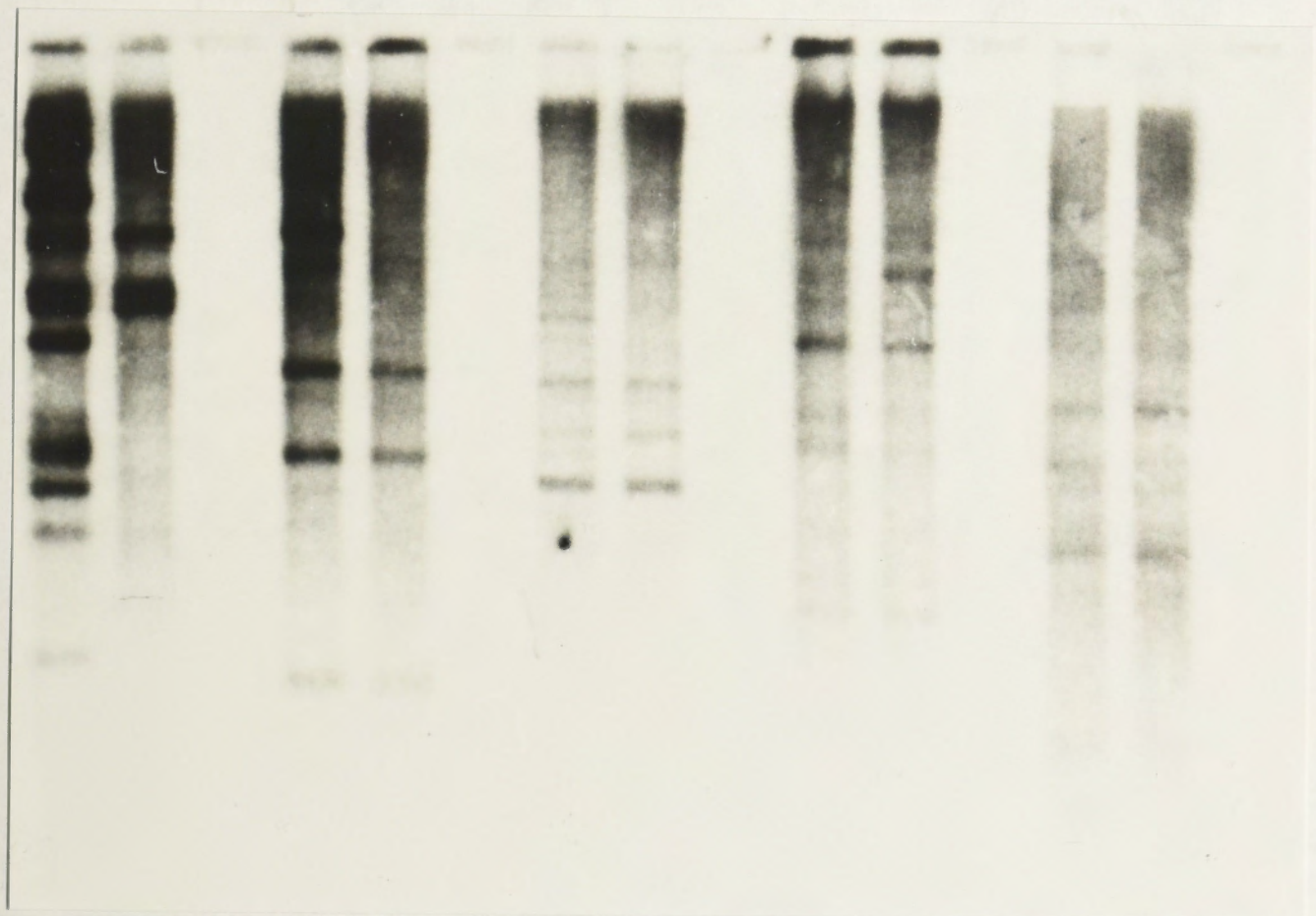


Figure 3.15. 2 µg of cattle male and female DNA (lanes 1 and 2), sheep male and female (lanes 3 and 4), goat male and female (lanes 5 and 6), deer male and female (lanes 7 and 8) and pig male and female (lanes 9 and 10) DNAs were digested with *Bam* HI and electrophoresed overnight at 12 volts in a 1% agarose gel, (**3.15A**). The DNAs were transferred by alkaline blotting to Zeta-Probe membrane and probed with nick-translated BRY.3, (**3.15B**). Lane 11 contains λ *Hind* III markers.

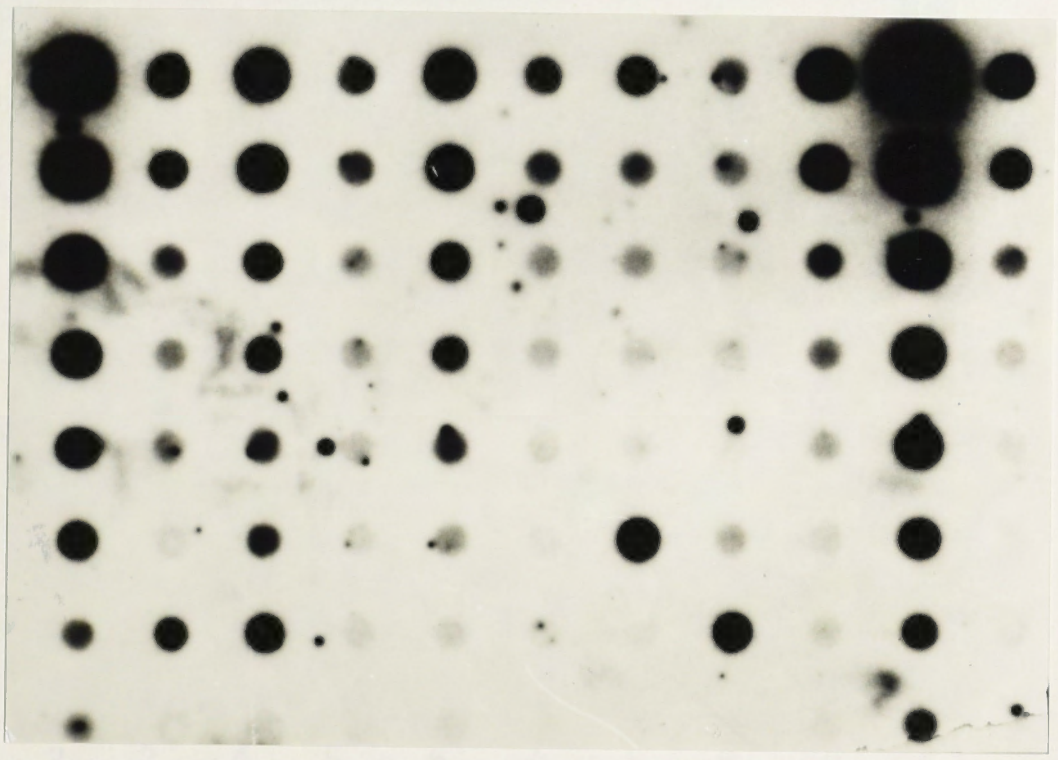


Figure 3.16. Genomic cattle male (lane 1) and female (2), sheep male (3) and female (4), goat male (5) and female (6) and deer male (7) and female (8) DNA was applied to a Zeta-Probe membrane in 0.4 N NaOH, using a Bio-Rad Dot Blot apparatus. The genomic DNAs were applied in doubling dilutions from 1 µg to 3.9 x 10^{-3} µg. Gel-purified insert from each of the subclones BRY.1 (9), BRY.2 (10)and BRY.3 (11) was also applied in doubling dilutions from 4 ng to 0.0156 ng. The blot was probed with the gel-purified insert from BRY.2, labelled by nick-translation.

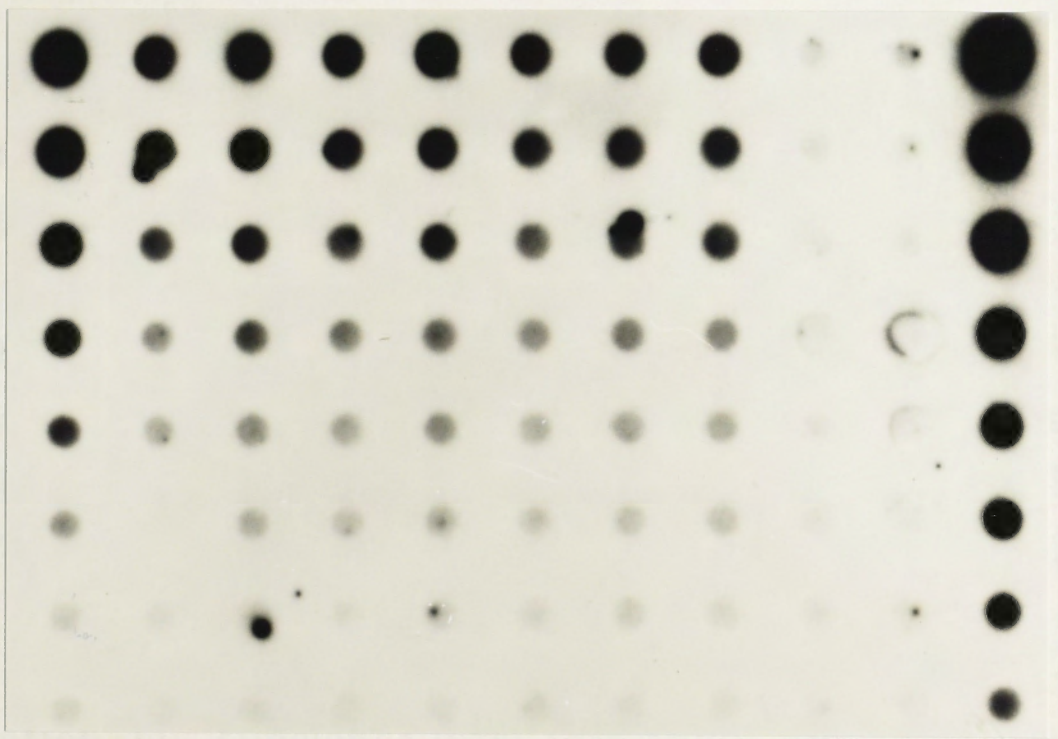


Figure 3.17. Genomic cattle male (lane 1) and female (2), sheep male (3) and female (4), goat male (5) and female (6) and deer male (7) and female (8) DNA was applied to a Zeta-Probe membrane in 0.4 N NaOH, using a Bio-Rad Dot Blot apparatus. The genomic DNAs were applied in doubling dilutions from 1 µg to 3.9×10^{-3} µg. Gel-purified insert from each of the subclones BRY.1 (9), BRY.2 (10)and BRY.3 (11) was also applied in doubling dilutions from 4 ng to 0.0156 ng. The blot was probed with the gel-purified insert from BRY.3, labelled by nick-translation.

While the copy number appears to decrease with the period of time the species have been separated from the line leading to cattle, this effect may simply be due to gradual sequence divergence, which has been very gradual considering the Bovidae and Cervidae diverged 30 million years ago. The dot blots were washed under stringent conditions, so that the level of homology required for hybridization is very high. Less stringent conditions may reveal many more sequences with slightly less homology to each of the probes.

3.4 Discussion

The isolation and characterization of a repeated *Sau* 3A1 fragment from the cattle Y chromosome (BRY.1) which has homology to repeated sequences on the Y chromosome of sheep and goats and to sequences found in both sexes of deer and pigs raised several questions concerning its relationship to other sequences on the Y chromosome and the reasons for its conservation. This fragment was used to isolate a 13 kb fragment of cattle DNA from a recombinant phage library. When the entire phage is used as a probe on Southern blots of genomic DNA there is no sex discrimination, but subclones from this phage hybridize to several male-specific restriction fragments. Most of this 13 kb (at least the 7.9 kb from which BRY.2 and BRY.3 are derived) does contain repeated sequences associated mainly or solely with the Y chromosome in cattle. These subclones also give a male-associated pattern of hybridization to sheep DNA and one of the subclones (BRY.2) also contains repeated sequences which are male-specific in goats and deer, showing that this 13 kb is from the Y chromosome. Within each subclone there is a group of related, repeated sequences, including BRY.1, which have elements shared between the subclones and elements found in only one of the subclones. Some of these sequences are repeated only on the Y chromosome, others are also repeated elsewhere in the genome. It is interesting to note that BRY.1 is probably not present in every group of sequences comprising BRY.2 and BRY.3 representatives since the copy number estimated for BRY.1 is so much lower (Chapter 2).

3.41 Genomic repeats on the bovine Y chromosome

Interspersed with the male-specific sequences in BRY.2 and BRY.3 there are repeated sequences which are present elsewhere in the genome

as well as on the Y chromosome. It would be interesting to investigate the question of whether these genomic repeats on the Y chromosome have homologues on the X chromosome because in this very stable chromosomal system (Wilson *et al.*, 1974) in which translocations are rare, the most likely source of additional sequences for the Y chromosome and the chromosome most likely to receive Y-chromosomal sequences is the X. The sex chromosomes of male eutherian mammals associate as a 'sex vesicle' during the early prophase of meiosis . Usually the axes of the X and Y pair at one end only to form a segment of synaptonemal complex which is indistinguishable from that of the autosomes (Solari, 1974). The pairing of homologous chromosomes or regions during meiosis is necessary both for disjunction of these elements at anaphase I and the exchange of material between them by crossing over. Homologous pairing usually involves synaptonemal complex formation (Moses, 1968), but not all synaptonemal complexes are between homologous regions (Gillies, 1974), and complex formation does not necessarily lead to crossing over. It could be postulated that there may be positive selection for the conservation of sequences involved in pairing of the X and Y chromosomes at meiosis.

The remaining 5.1 kb of the phage insert (the regions outside BRY.2 and BRY.3) contain sequences which are also found elsewhere in the genome in sufficient numbers to totally obscure the male-specific pattern obtained with the subcloned fragments. These highly repeated sequences are not representatives of the 5 major classes of cattle satellite DNAs which are absent from the sex chromosomes of cattle (Kurnit *et al.*, 1978). This is confirmed by the fact that the Y-chromosomal sequences from EMBL3A.Y1 which flank BRY.2 and BRY.3 hybridize strongly to sheep genomic DNA (Figures 3.7 and 3.8). While the two sheep and goat satellites have been conserved and show close DNA homology, they do not have homology with any of the cattle satellites (Kurnit *et al.*, 1978). The further information that at least the sequences contained within the 2.1 kb of Y-chromosomal DNA attached to the right arm of the phage do not hybridize to genomic goat DNA finally excludes the possibility that these sequences belong to one of the satellite classes.

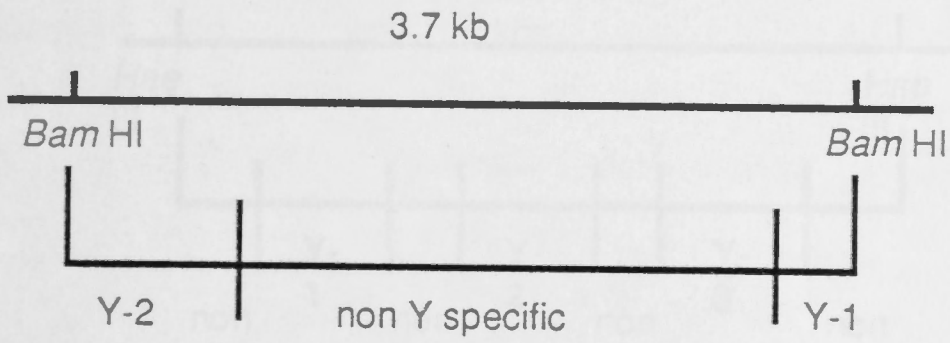
It is possible these sequences represent Y-chromosomal copies of the artiodactyl repeats (Watanabe *et al.*, 1982). These short interspersed repeats (SINEs; Singer, 1982) are about 720 bp long and there are about 10^5

copies in the bovine genome. They show similarities to the human *Alu* repeats (Jelinek *et al.*, 1979), the mouse B family (Krayev *et al.*, 1980), rat sequences (Page *et al.*, 1981) and the Chinese hamster *Alu* equivalent family (Haynes and Jelinek, 1981). Most of these repetitive sequences show two common features: (i) a dA-rich sequence present at their 3' ends, and (ii) direct repeats of a short non-conserved sequence flanking each end. For the bovine repetitive sequence no dA-rich sequence is found at the 3' end; however, there are tandem repeats of $(AGC)_x$ or $(CACT_{y)z}$. These repetitive sequences are bounded by non-conserved short direct repeats, implying their presence may be the result of transposition events. Some of the interspersed repetitive DNA sequences are transcribed into low molecular weight RNAs. Like the *Alu* repeats in the human genome, these sequences seem to achieve their wide dispersal via mechanisms that involve RNA intermediates (Jelinek *et al.*, 1979). Although no members of the artiodactyl repeat family were found within BRY.2 or BRY.3 (Chapter 4), they have been found in sequences from the bovine Y chromosome (K. Matthaei, personal communication), and sheep Y-chromosomal homologues of BRY.2 (other than OY1) contain representatives of these sequences (Lord, 1989).

3.42 The structure of the bovine Y chromosome

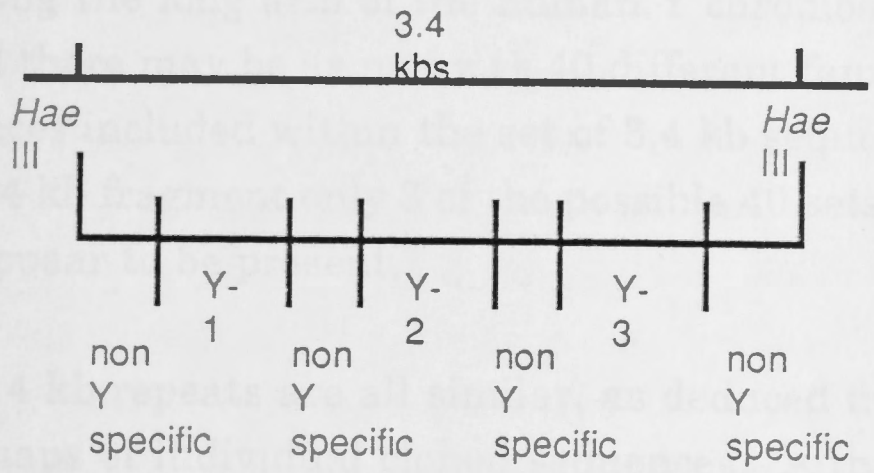
The organization of sequences within BRY.2 as determined by hybridization of restriction fragments from BRY.2 with Southern blots of genomic cattle DNAs and the probing of such fragments with BRY.1 is shown diagrammatically in Fig.3.18. The relationships of Y-specific repeated sequences and non-Y-specific repeats is shown and can be compared with the molecular structure of the human Y chromosome proposed by Kunkel and Smith (1982), shown in Figure 3.19. There are several striking similarities between the organization of repeats on the cattle and human Y chromosomes, including the presence of several different Y-specific repeated sequences and their interspersion with sequences which are not confined to the Y chromosome.

The human Y chromosome was found to be composed of a variety of repeated DNAs, including the alphoid repeats, the *Alu* repeats and several long interspersed repeats (LINEs, Singer,1982) which could not be distinguished from other human LINE repeats (Smith *et al.*, 1987). The human Y chromosome also contains a group of repeated DNA elements



Scale: 2 cm = 1 kb

Figure 3.18. Schematic diagram of BRY.2 for comparison with the human Y chromosomal 3.4 kb *Hae* III repeats (**Figure 3.19**). The Y-specific repeats (Y1=BRY.1, Y2= BRY.2A) are interspersed with non-Y-specific repeats.



Scale: 2 cm = 1 kb.

Figure 3.19. Schematic representation of the human male 3.4 kb family generated by restriction with *Hae* III, from Kunkel and Smith (1982). Each fragment has sequences that are specific to the Y chromosome, but may differ from one another (Y-1, Y-2, Y-3), as well as sequences that occur on other chromosomes (non Y specific). The actual arrangement of these components is not entirely known, but compare this postulated organization with the organization of repeats in the cattle Ychromosomal DNA in BRY.2 (**Figure 3.18**).

originally identified as 3.4 and 2.1 kb fragments in *Hae* III digests of male genomic DNA (Cooke, 1976). The 3.4 and 2.1 kb fragments do not cross-react (Kunkel and Smith, 1982), but both exist as tandem clusters of alternating Y-specific and non-Y-specific sequences (Smith *et al.*, 1987). The 3.4 kb human Y-chromosomal repeats contain at least three distinct sequences with autosomal homologies (Kunkel and Smith, 1982), interspersed in various ways with a collection of several different Y-specific repeat sequences. Individual clones derived from isolated 3.4 kb *Hae* III fragments have been identified which do not cross react therefore the 3.4 kb *Hae* III fragments are a heterogeneous mixture of sequences which have in common the regular occurrence of *Hae* III restriction sites at 3.4 kb intervals along the long arm of the human Y chromosome. Kunkel *et al* (1976) found there may be as many as 40 different families of Y specific DNA sequences included within the set of 3.4 kb sequences, yet within any particular 3.4 kb fragment only 3 of the possible 40 sets of reiterated sequences appear to be present.

The 2.4 kb repeats are all similar, as deduced from the comparison of restriction maps of individual cloned sequences. Although these Y-chromosomal repeats are tandemly arrayed they are distributed along the length of the long arm and account for only 25% of Y long arm DNA (Cooke, 1976). They must therefore occur as independent tandem clusters interspersed among other sequences rather than as a single tandem array.

Since there are at least superficial similarities between the organization of repeat structures of the human and cattle Y chromosomes, though the human repeats are not Y-specific in closely related species, it may be worthwhile examining the theories suggested for the origin of the pattern of the human Y chromosome. The major human Y-chromosomal tandem repeats are not known to be transcribed so it is unlikely that they have become associated with the the human Y chromosome by a mechanism involving an RNA intermediate. It is more likely that a series of translocations and/or transpositions involving the X and several autosomes has occurred, especially since these sequences are on the X and/or autosomes but not the Y chromosome of other primates (Smith *et al.*, 1987). These multiple transposition events could have been followed by selective amplification of particular sequences resulting in a variety of Y-chromosomal tandem repeats. The heterogeneity of sequences within

each family of tandem repeats suggests such amplification events must have occurred several times. Repeats may have been generated as single tandem arrays then shuffled to various sites along the Y long arm, perhaps as a result of an inherent instability of the non-pairing region of the Y chromosome.

There is no evidence in the hybridization pattern of BRY.2 or BRY.3 to *Bam* HI restriction digests of genomic DNA for a tandem repeat, and the sequence data (Chapter 4) does not show a discernible repeat unit, however the insert of cattle DNA in EMBL3A.Y1 may be only part of a much longer repeat unit on the Y chromosome.

3.43 Evolution of the artiodactyl Y chromosome

A significant proportion of the cattle Y chromosome (40%) is made up of related DNA sequences which are confined mainly to the Y chromosome. Most of these sequences also have homologues on the Y chromosome of closely and not so closely related species and these homologues are confined mostly to the Y chromosome in these species as well, so that conservation of repeats on the Y chromosome of sheep, cattle, goats and deer is much greater than first thought after the isolation of BRY.1. This suggests that a large proportion of the Y-specific sequences on the Y chromosome of the Bovid and Cervid families have been substantially unchanged since the divergence of these groups from other artiodactyls, and that the amplification of these sequences on the Y chromosome must have occurred soon after this divergence, and before the divergence of cattle, sheep and goats (Bovidae) from deer (Cervidae) 30 million years ago (Goodman *et al.*, 1982). This is particularly interesting because BRY.1 is not repeated in deer and does not discriminate between the sexes in this species (Chapter 2), though it is part of BRY.2 which is repeated and male-enriched in deer. This may reflect the fact that BRY.1 is present at a much lower copy number in cattle and therefore probably is not included in every BRY.2 homologue. The lack of amplification in deer may indicate that BRY.1 has become incorporated into the BRY.2 group of sequences after the divergence of the Bovidae from the Cervidae.

The absence of male specificity of BRY.3 in goats and deer suggests a more recent (less than 5 million years) spread to the autosomes or perhaps an amplification of a few representative sequences which were already on

the autosomes and/or X. It is also possible that BRY.3 (which hybridizes more strongly to female DNA than does BRY.2; Figures 3.14 and 3.15) contains copies of genomic repeats which have greater numbers of autosomal representatives in the genomes of goats and deer than in cattle and sheep. Other repeated sequences which are found in the genomes of both bovids and cervids have been described. One of these is an endogenous type C virus isolated from cells of the Columbian black-tailed deer *Odocoileus hemionus* (Aaronson *et al.*, 1976) which is also found in the Bovidae, suggesting the genetic transmission of type C viral genes within cervids and bovids for at least 25-30 million years.

The results presented in this Chapter suggest that the Y-specific repeated sequences conserved on the cattle, sheep, goat and deer Y chromosomes have originated from a very early translocation from the X and/or autosomes in a progenitor, then undergone amplification event(s) increasing copy number. McKay *et al.* (1978) cite the fact that a DNA sequence can vary in amount without a concurrent alteration in the sequence itself as evidence for saltatory replication, that is in a few large steps. This process, perhaps via intra-chromosomal exchange (Smith, 1976), may have been responsible for the original amplification of these sequences on the progenitor Y chromosome. The fact that the repeated sequences are not confined to the Y chromosome of all artiodactyls, combined with differing copy numbers in the different species indicates that conservation of the sequence is not sequence-dependent, that is not due to some intrinsic function of the sequence itself. The repeats may also be spread along the Y chromosome rather than confined to one region, suggesting further that their conservation is not due to a close association with coding sequences. Perhaps these sequences were isolated from recombination with the rest of the genome because of the special role of the Y chromosome in sex determination, so were not lost or spread to other chromosomes and their conservation may be due solely to the genetic isolation of the Y chromosome

The interspersion of non-Y-specific repeated sequences with Y-specific repeated sequences on the cattle Y chromosome indicates it has not been genetically isolated for all of the period since it acquired these repeats from the autosomes. It is possible that the cattle Y chromosome, may once have had a tandem repeat structure, but that during the much

longer time since the original amplification events (in comparison to man) so much scrambling and interspersion of other sequences has occurred that the tandem blocks have been broken up. It has been suggested (Watson *et al.*, 1987) that tandem repeats of an identical DNA sequence would be an ideal substrate for intra-chromosomal recombination, which would lead to loss of tandemly repeated families.

The Y chromosome is not immune to colonization by sequences which spread by RNA intermediates, so the originally tandem Y-specific repeated sequences may have been interspersed with the *Alu*-like artiodactyl repeats and perhaps the other non-Y-specific sequences found within BRY.2 and BRY.3 are also sequences which do not rely on conventional methods of recombination for their dispersal. The retroposition of non-viral cellular RNA species has emerged as a major evolutionary force, contributing to continuous sequence duplication, dispersion and rearrangement of eukaryotic genomes (Weiner *et al.*, 1986). The two subclones were sequenced to obtain more information about the short term organization of the regions which contain Y-specific and non-Y-specific sequences and the results are presented in Chapter 4.

CHAPTER FOUR

SEQUENCE ANALYSIS OF Y-CHROMOSOMAL SEQUENCES

CHAPTER FOUR

Sequence Analysis of Y-Chromosomal Sequences

4.1 Introduction

The isolation and characterization of EMBL3A.Y1 showed that the Y-specific repeated sequence BRY.1 is present in a block of sequences on the cattle Y chromosome which also contains other, different Y-specific repeated sequences and repeated sequences which are not confined to the Y chromosome. BRY.1 hybridizes to both of the *Bam* HI fragments subcloned from this phage and these two fragments produce a male-associated pattern of hybridization to Southern blots of genomic DNA which is not seen with the whole phage or the cattle DNA either side of BRY.2 and BRY.3 in this recombinant phage.

The subclones were sequenced and the data was analysed for the presence of known artiodactyl SINEs. The sequences were searched for signs of RNA processing, such as poly(A) 'tails', direct repeats and possible open reading frames which may indicate retroposition of an mRNA to produce a pseudogene, and for inverted repeats which may indicate the insertion of retroviral sequences (Rogers, 1985). LINEs also contain signs of RNA processing but they are long sequences with little evidence of internal repetition (Adams *et al.*, 1980), so the lack of long direct repeats may indicate these sequences are similar to the characterised LINE families.

Sheep Y-chromosomal homologues to BRY.2 had been isolated and sequenced (Lord, 1989) and it was of interest to compare the data for the two species to determine the level of homology between similar regions and to study the differences in the short-term structure of these sequences on the cattle and sheep Y chromosomes.

4.2 Materials and Methods

4.21 Sequencing

The *Bam* HI fragments from EMBL3A.Y1 which had been subcloned into pTZ18U and pTZ19U (Chapter 3) were sequenced. The recovery of each

of the fragments in both orientations was confirmed by digestion with enzymes found to cut the insert asymmetrically, *Hind* III for BRY.2 and *Eco* RI for BRY.3. A series of unidirectional over-lapping deletion clones was obtained for both the sequenced fragments in each direction using the exonuclease III method (Henikoff, 1984) and the Erase-a-base kit supplied by Promega. For each of the subclones suitable enzymes were selected to provide a 5' overhang for digestion of the insert by exonuclease III after restriction with an appropriate enzyme to provide a 3' overhang on the primer side of the vector which affords protection from the exonuclease. Where suitable enzymes were not available for digestion to provide a protective 3' overhang (due to the presence of a site within the insert), an enzyme generating a 5' overhang was used then end-filled with α-phosphorothioate deoxynucleotides and Klenow fragment of DNA polymerase (Putney *et al.*, 1981). Both subclones were sequenced in both directions.

Single-stranded DNA (ssDNA) templates were prepared and sequenced as described (Chapter 2).

4.22 Sequence analysis

The sequencing autoradiographs were read using digitizing hardware and software provided by the Gene-Master system (Bio-Rad). Sequences were analysed for the presence of open reading frames (ORFs), polyadenylation sites, inverted repeats and direct repeats. Comparison of sequences was done using the DIAGON dot matrix programs (Staden, 1982) and regions showing homology were aligned using the Needleman-Wunsch alignment (Needleman and Wunsch, 1970). The DNA Inspector IIe (Version 2) program for the Macintosh computer was also used for sequence analysis.

The nucleotide sequences of BRY.2 and BRY.3 were compared with each other and with BRY.1 (Chapter 2). They were also compared with OY.1, a 5,195 bp sheep Y-chromosomal sequence isolated using BRY.2 as a probe (Lord, 1989), with BRY4CART, a 293 bp representative of an artiodactyl SINE family isolated by K. Matthaei from a cattle Y-chromosomal fragment and with a 319 bp cDNA clone from an adult bull testis cDNA library which contains another member of this family, isolated by S. Beaton. The cattle sequences were also compared with OY.11.1, another sheep Y-chromosomal sequence subcloned from a recombinant

phage isolated using BRY.2 (Lord, 1989). This sequence is 3,980 bp long, contains an open reading frame and has homology with a cDNA clone from an adult bull testis cDNA library (S. Beaton, personal communication).

4.3 Results

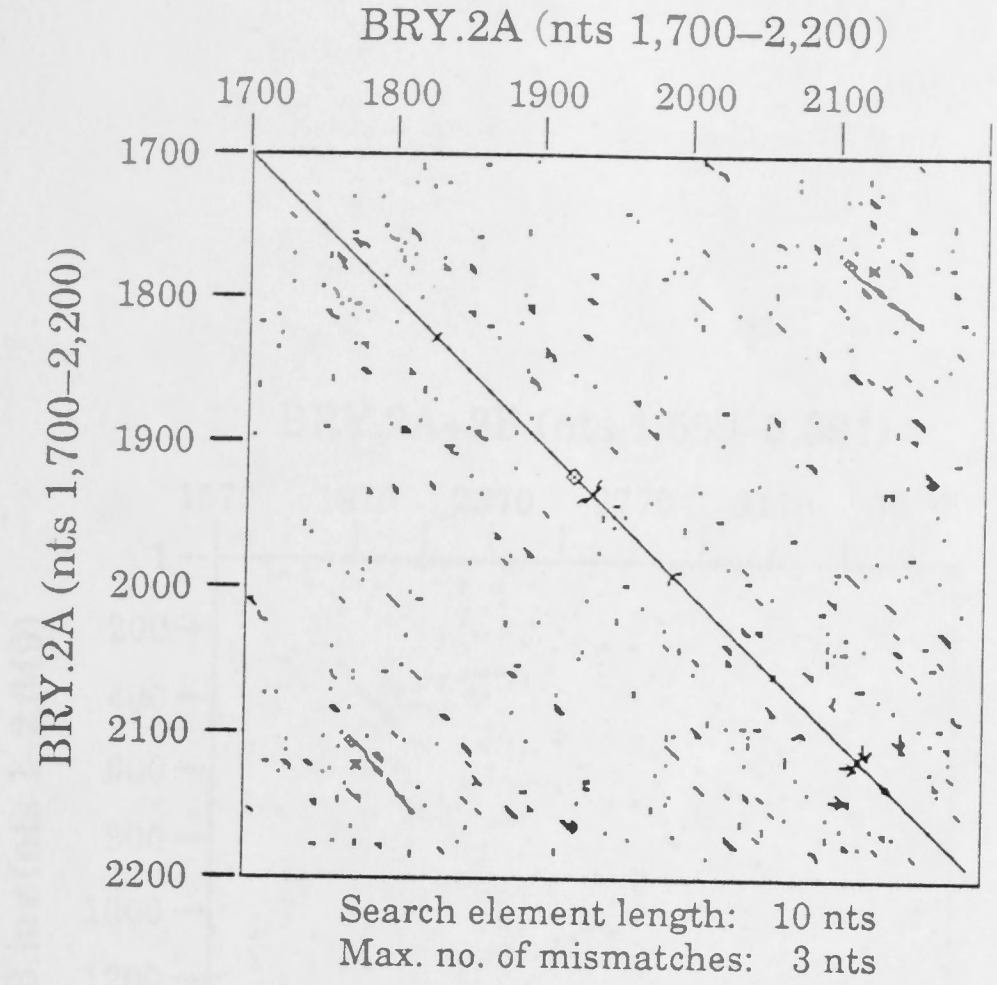
4.31 BRY.2 and BRY.3 sequence information

The sequence data for BRY.2 and BRY.3 is shown in Appendices 2 and 3. The sequences are presented 5' to 3' relative to their orientation in EMBL3A.Y1 (Chapter 3). BRY.2 is 3,692 nucleotides long. Only partial sequence data for BRY.3 was obtained. Despite repeated and frustrating attempts, sequence data for ss templates corresponding to the region from 1,359 bp to 2,188 bp could not be determined. The sequence which has been determined is presented in Appendix 3 as two different regions: BRY.3A (1,359 bp) and BRY.3B (2,012 bp).

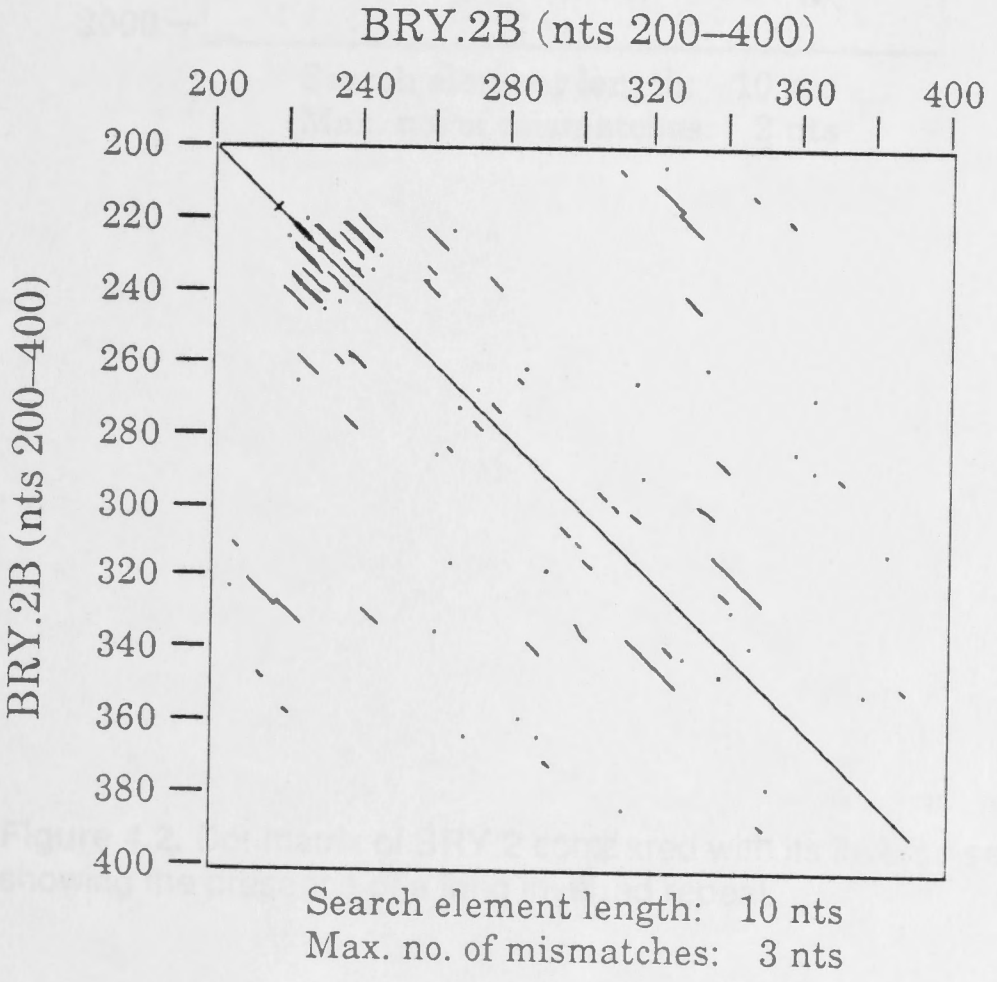
The sequences were examined for the presence of internal repeats, including alternating purine-pyrimidine tracts. BRY.2 was compared as two fragments: BRY.2A which corresponds to nt 1-2,666 in Appendix 2 and BRY.2B which corresponds to nt 2,659-3,692.

BRY.2 contains a 52 bp direct repeat which begins at nucleotide 1,170 and at nucleotide 2,119, with 11 mismatches between the two copies. The dot matrix comparison showing the presence of this long direct repeat is shown in Figure 4.1A and it is underlined in the sequence in Appendix 2. Several short direct repeats are also apparent from this dot matrix. Figure 4.1B shows that BRY.2 contains a great many short direct repeats. A search identified 13 pairs of perfect direct repeats of 9, 10 or 11 bp in this region of BRY.2. BRY.2 also contains a long inverted repeat which can be seen in Figure 4.2. This sequence is 131 nucleotides long (first copy begins at nt 1,965, second copy begins at nt 2,119) and contains only 11 mismatched bases. The dot matrices in Figure 4.3 show an expansion of the regions containing these inverted repeats.

BRY.2 was found to contain an imperfect (TG).$(AC)_{11}$ tract which is underlined in Appendix 2.



4.1A



4.1B

Figure 4.1. Dot matrices showing the presence of direct repeats within BRY.2. BRY.2A (nucleotides 1- 2666 in Appendix 2) contains a long direst repeat (**4.1A**), while BRY.2B (nucleotides 2559-3692) contains a large number of short perfect direct repeats (**4.1B**).

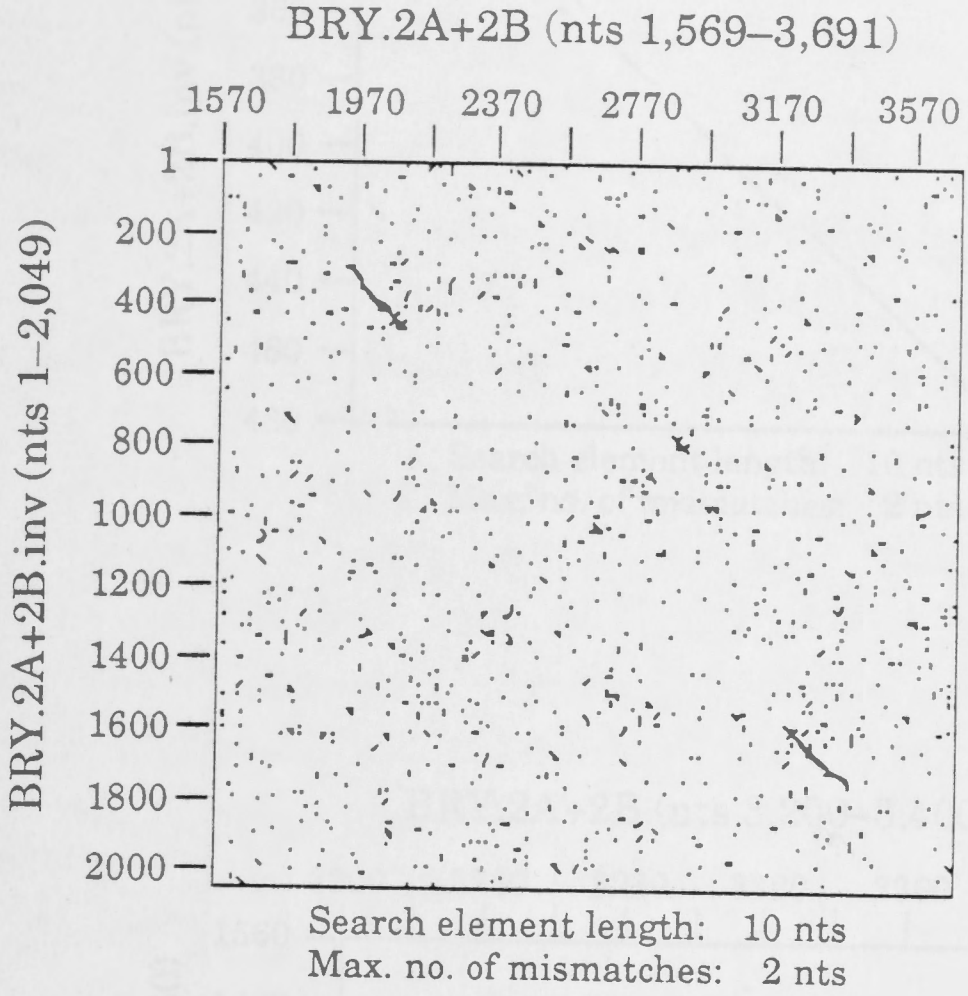


Figure 4.2. Dot matrix of BRY.2 compared with its inverse sequence, showing the presence of a long inverted repeat.

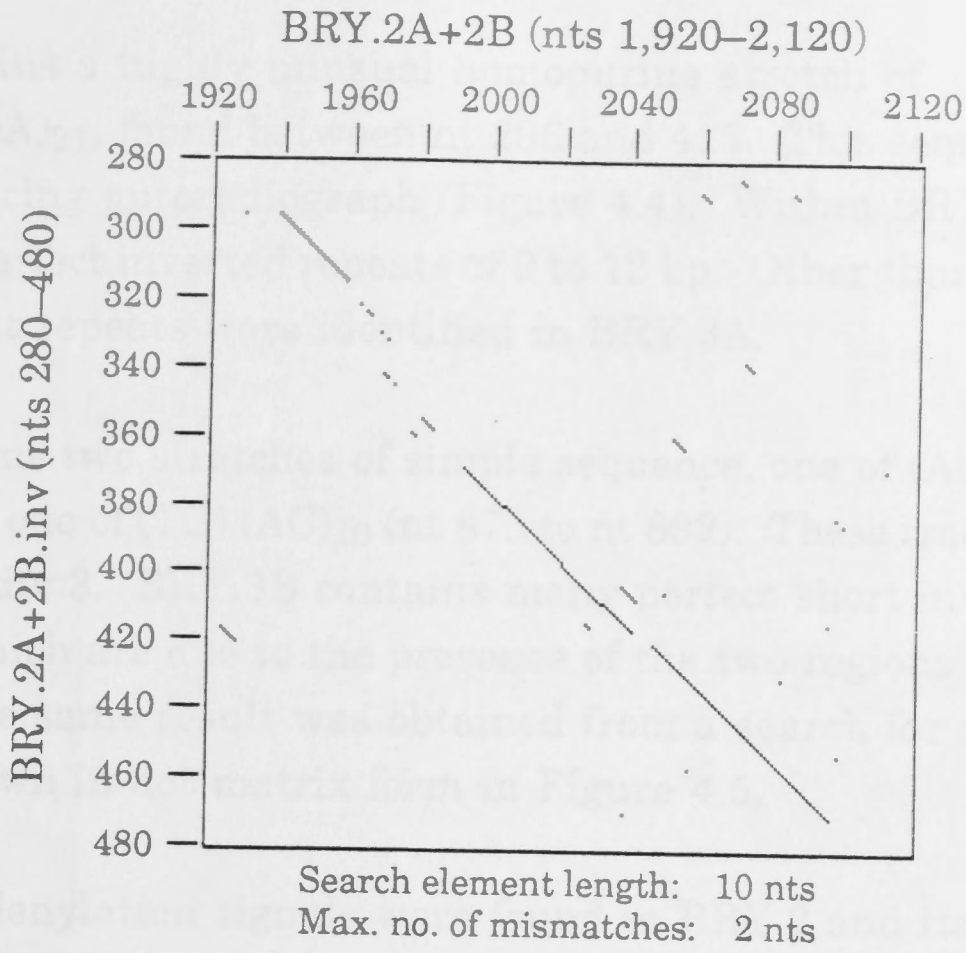


Figure 4.3. Dot matrices of BRY.2 compared with its inverse sequence, showing the 1st (**4.3A**) and 2nd (**4.3B**) copies of the long inverted repeat seen in Figure 4.2. In each case an area of Figure 4.2 has been expanded.

BRY.3A contains a highly unusual homopurine stretch of $(GA)_{24}.(GGAGA)_6.(GA)_{21}$, found between nt 296 and 415. This sequence is shown in the sequencing autoradiograph (Figure 4.4). Within BRY.3A there are 6 pairs of perfect inverted repeats of 9 to 12 bp. Other than the (GA) stretch no direct repeats were identified in BRY.3A.

BRY.3B contains two stretches of simple sequence, one of $(AC).(TG)_{15}$ (nt 674 to nt 794) and one of $(TG).(AC)_{10}$ (nt 873 to nt 892). These tracts are underlined in Appendix 3. BRY.3B contains many perfect short inverted repeats, several of which are due to the presence of the two regions of simple sequence. The same result was obtained from a search for direct repeats, which is shown in dot matrix form in Figure 4.5.

Possible polyadenylation signals were found in BRY.2 and its inverse (reversed and complemented) sequence, but none were identified in BRY.3. Open reading frames (ORFs) were identified in BRY.2 and in both the sequenced regions of BRY.3. Several of these ORFs are of a reasonable length, i.e. longer than 300 nts: BRY.2 (324: 719-1043 and 393: 3260-3653). The inverse of BRY.2 contains long ORFs in two frames: (345: nt 375-720 and 414: nt 779-1193). In BRY.3B there are two ORFs, one of 351 nt from nt 39-390, the other of 405 nt from nt 844-1249. The inverse sequence of BRY.3B contains one in each possible frame, one of 429 nt from nt 966-1,395, a second of 312 nt from nt 1484-1796 and a third of 315 nt from nt 1108-1423.

The G plus C contents of BRY.2, BRY.3A and BRY.3B are presented in Table 4.1.

Except for the stretches of simple sequence no sequence similarities were found between BRY.2 and BRY.3. No homology was found between the two sequenced regions of BRY.3.

BRY.1 was found to be in the region predicted by hybridization to Southern blots of BRY.2 (Chapter 3). This region of sequence homology is underlined in Appendix 2 and is shown in dot matrix form in Figure 4.6. The BRY.1 sequence begins at nt 3,380 in BRY.2 and continues to the *Bam* HI (*Sau* 3AI) cloning site. Six bases present in the original copy of BRY.1 are absent from BRY.2, giving an overall identity of 98% as shown by Needleman-Wunsch alignment. BRY.1 does not have sequence identity

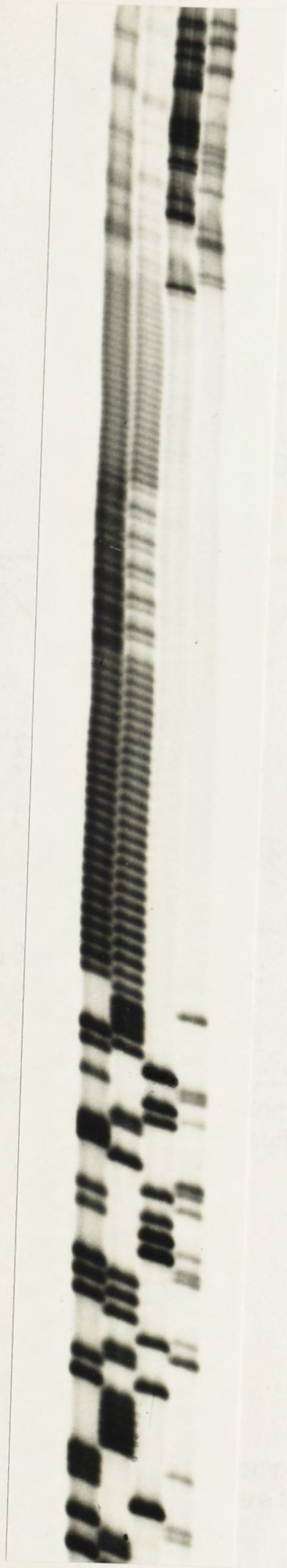


Figure 4.4. A section from a sequencing gel showing a homopurine tract, $(GA)_{24}$, $(GGAGA)_6$, $(GA)_{21}$ in BRY.3A.

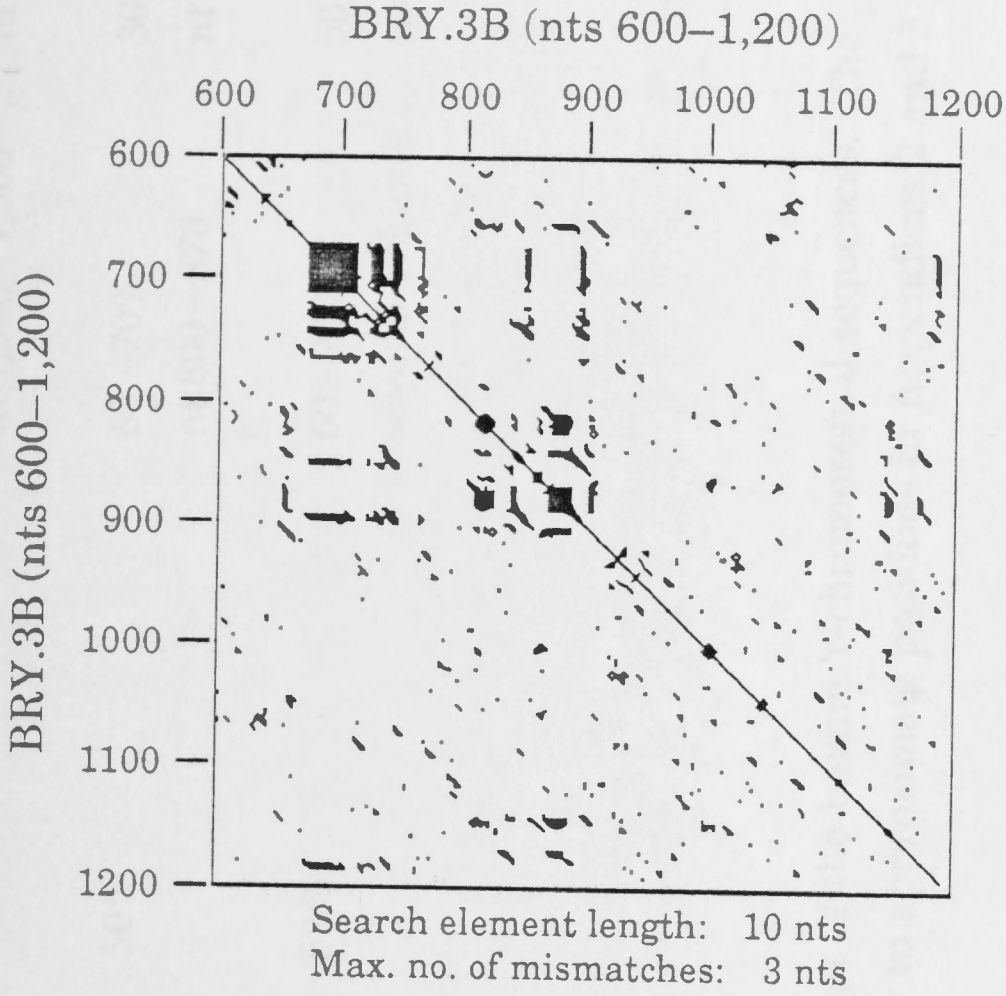
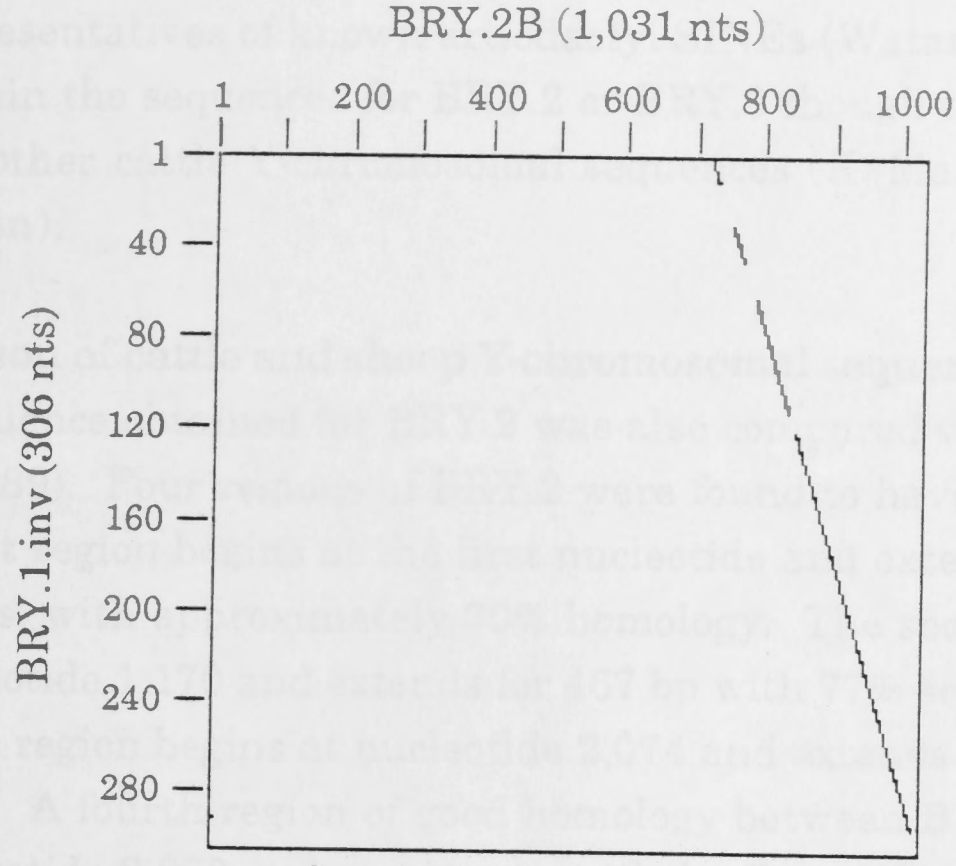


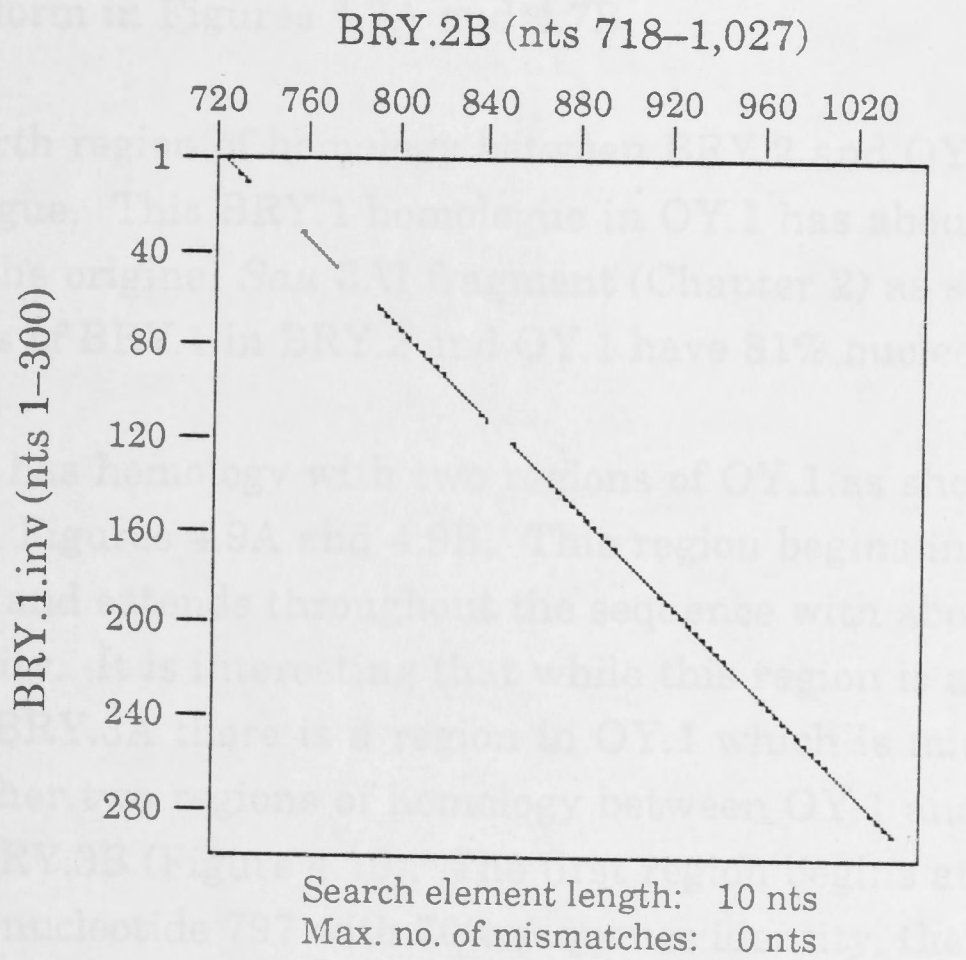
Figure 4.5. Dot matrix showing a comparison between BRY.3B and itself. The only direct repeats are the stretches of simple sequence.

	G and C content	Highest range	Lowest range
BRY.2	52%	60 - 70% nt 1,300 - 1,500	30 - 40% nt 2,800 - 3,000
BRY.3A	50%	60 - 70% nt 800 - 870	30 - 40% nt 470 - 670
BRY.3B	52%	60 - 70% several regions	50 - 60% several regions

Table 4.1. G and C content of the bovine Y-chromosomal sequences. The nucleotide numbers refer to approximate positions in Appendices 2 and 3.



4.6A



4.6B

Figure 4.6. Dot matrices of BRY.2B compared with BRY.1, showing the region of BRY.2 in which BRY.1 is found (**4.6A**) and an expansion of this region of homology (**4.6B**). Nucleotide 720 in BRY.2B is equivalent to nucleotide 3380 inAppendix 2.

with any of the 3,371 bp of sequence obtained for BRY.3. However, it does hybridize with BRY.3 (Chapter 3) so this sequence must be assumed to be part of the missing region of sequence.

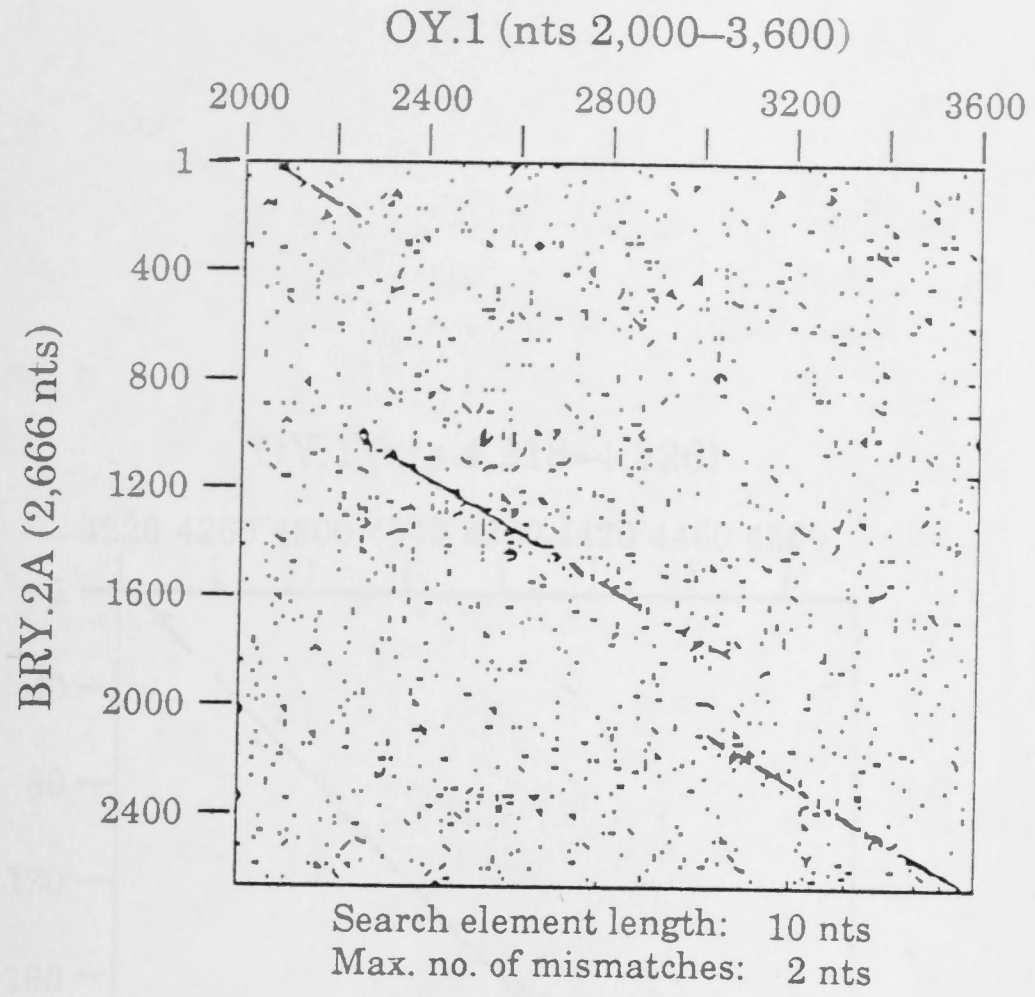
No representatives of known artiodactyl SINEs (Watanabe, 1982) were identified within the sequences for BRY.2 or BRY.3 though they have been found within other cattle Y-chromosomal sequences (K. Matthaei, personal communication).

4.32 Comparison of cattle and sheep Y-chromosomal sequences

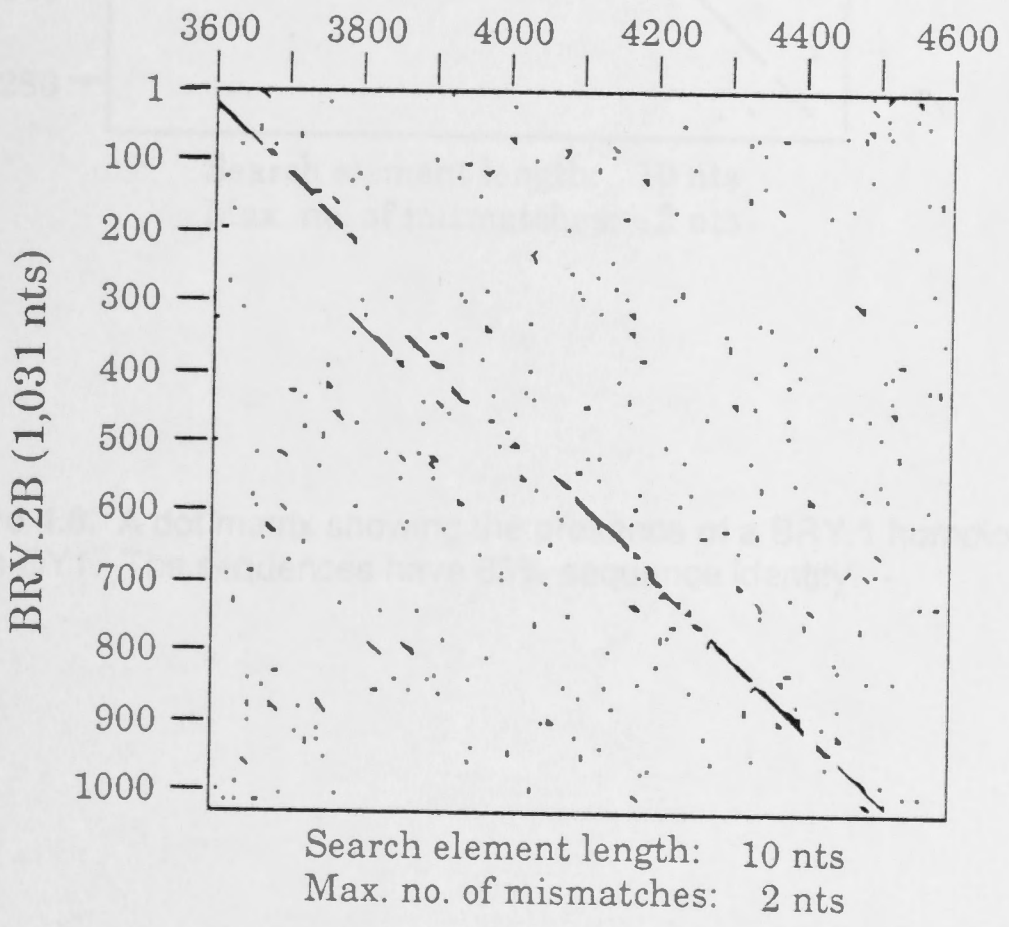
The sequence obtained for BRY.2 was also compared with that of OY.1 (Lord, 1989). Four regions of BRY.2 were found to have homology with OY.1. The first region begins at the first nucleotide and extends for about 200 nucleotides, with approximately 70% homology. The second region begins at nucleotide 1,170 and extends for 467 bp with 77% sequence identity, while the third region begins at nucleotide 2,074 and extends for 704 bp with 62% homology. A fourth region of good homology between BRY.2 and OY.1 begins at nucleotide 2,970 and continues until the *Bam* HI cloning site with 78% homology. These regions are underlined in Appendix 2 and presented in dot matrix form in Figures 4.7A and 4.7B.

The fourth region of homology between BRY.2 and OY.1 includes the BRY.1 homologue. This BRY.1 homologue in OY.1 has about 83% sequence identity with the original *Sau* 3AI fragment (Chapter 2) as shown in Figure 4.8. The copies of BRY.1 in BRY.2 and OY.1 have 81% nucleotide identity.

BRY.3A has homology with two regions of OY.1 as shown in dot matrix form in Figures 4.9A and 4.9B. This region begins in BRY.3A at nucleotide 804 and extends throughout the sequence with about 70% sequence identity. It is interesting that while this region is almost continuous in BRY.3A there is a region in OY.1 which is missing from BRY.3A. Another two regions of homology between OY.1 and BRY.3 were found within BRY.3B (Figure 4.10). The first region begins at nucleotide 10 and extends to nucleotide 797 with 70% sequence identity, then there is a short break in homology and it resumes again at nucleotide 1,015 and continues until the end of the sequence with about 73% sequence identity.



4.7A



4.7B

Figure 4.7. Dot matrices showing comparisons of the sequences of of BRY.2 with OY1. **4.7A** is an expansion of the dot matrix obtained for the first region of BRY.2 against OY1, while **4.7B** is an expansion of the homologous region in BRY.2B. Nucleotide 3600 in BRY.2B is equilavent with nucleotide 2559 in Appendix 2.

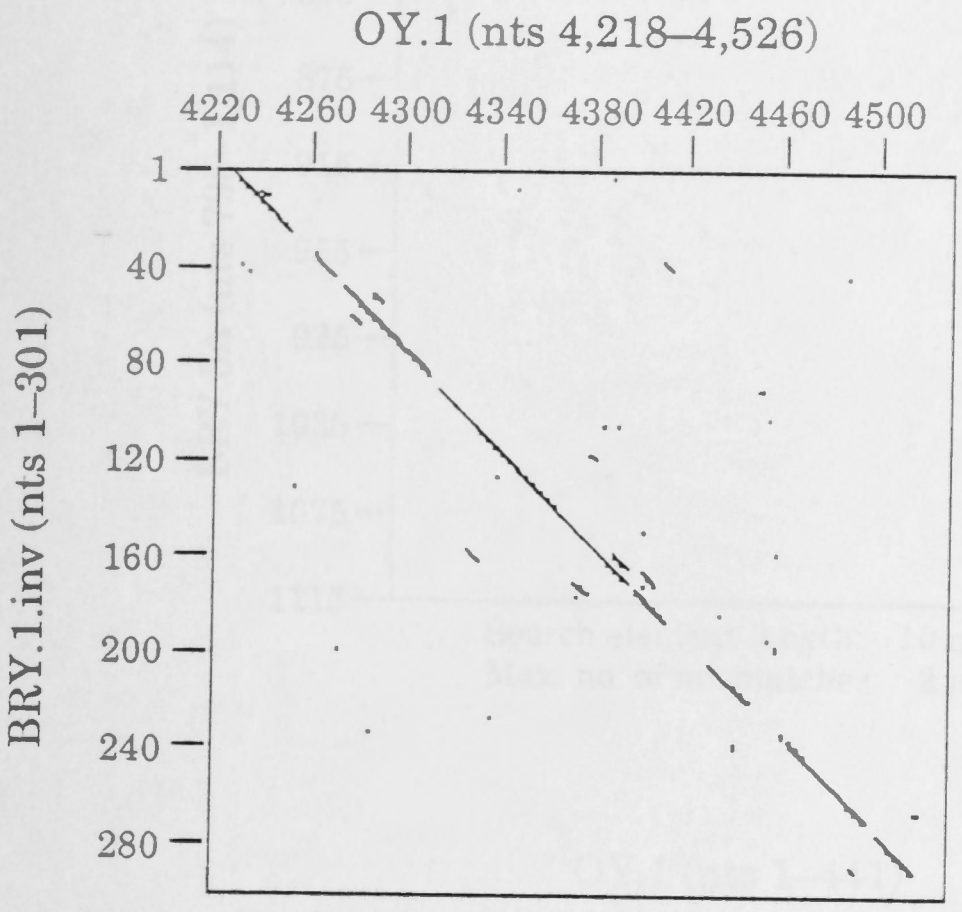
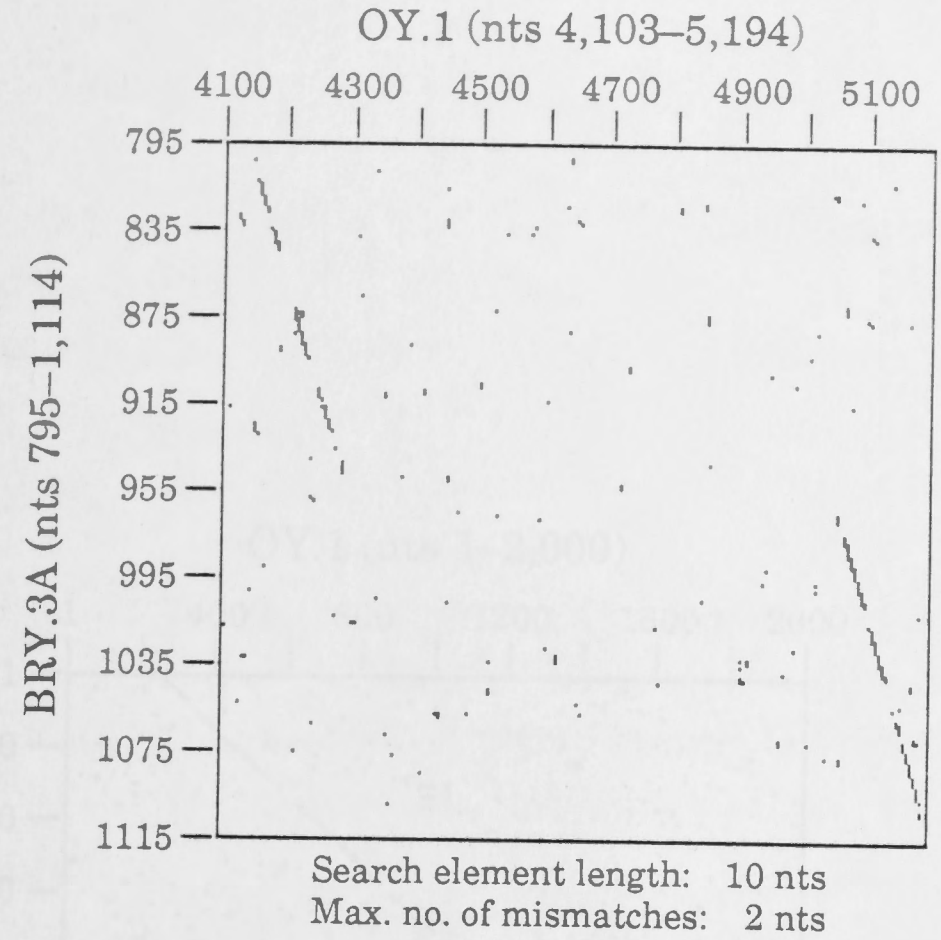
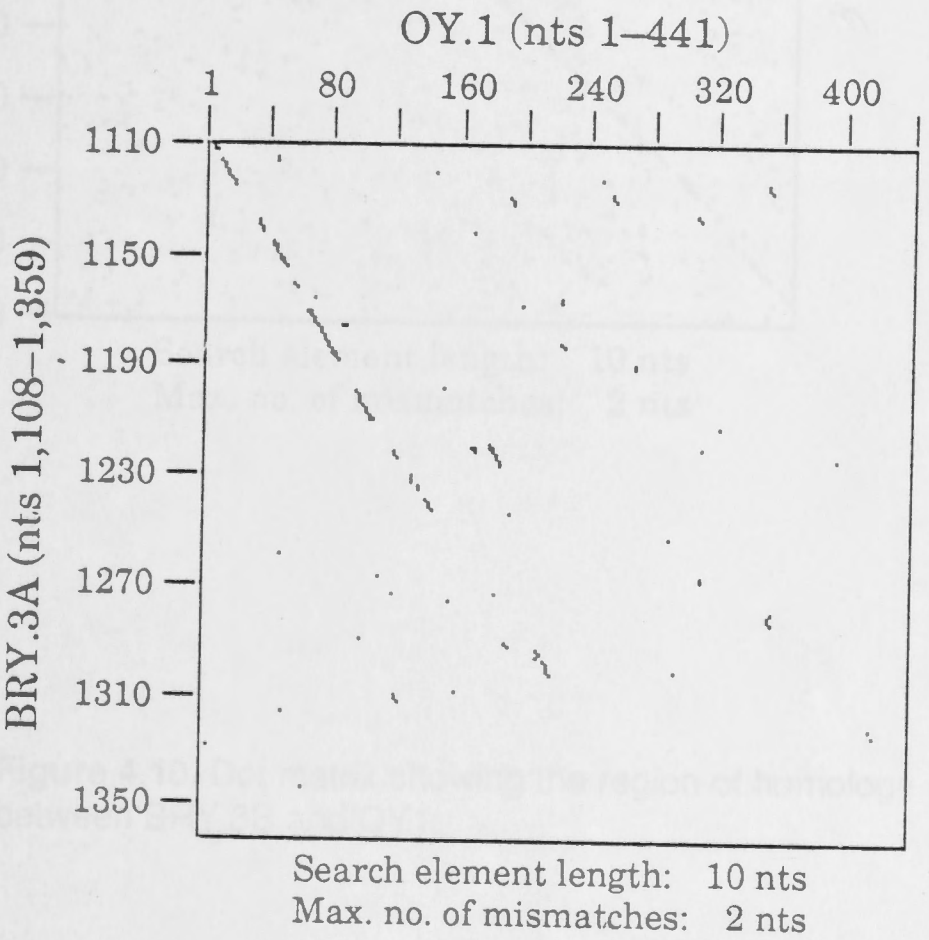


Figure 4.8. A dot matrix showing the presence of a BRY.1 homologue within OY1. The sequences have 83% sequence identity.



4.9A



4.9B

Figure 4.9. Dot matrices showing the regions of sequence homology between BRY.3A and OY1. It can be seen that the region shown in **4.9A** continues throughout the sequence (**4.9B**). A deletion in OY1 relative to BRY.3A can be seen in **4.9B**.

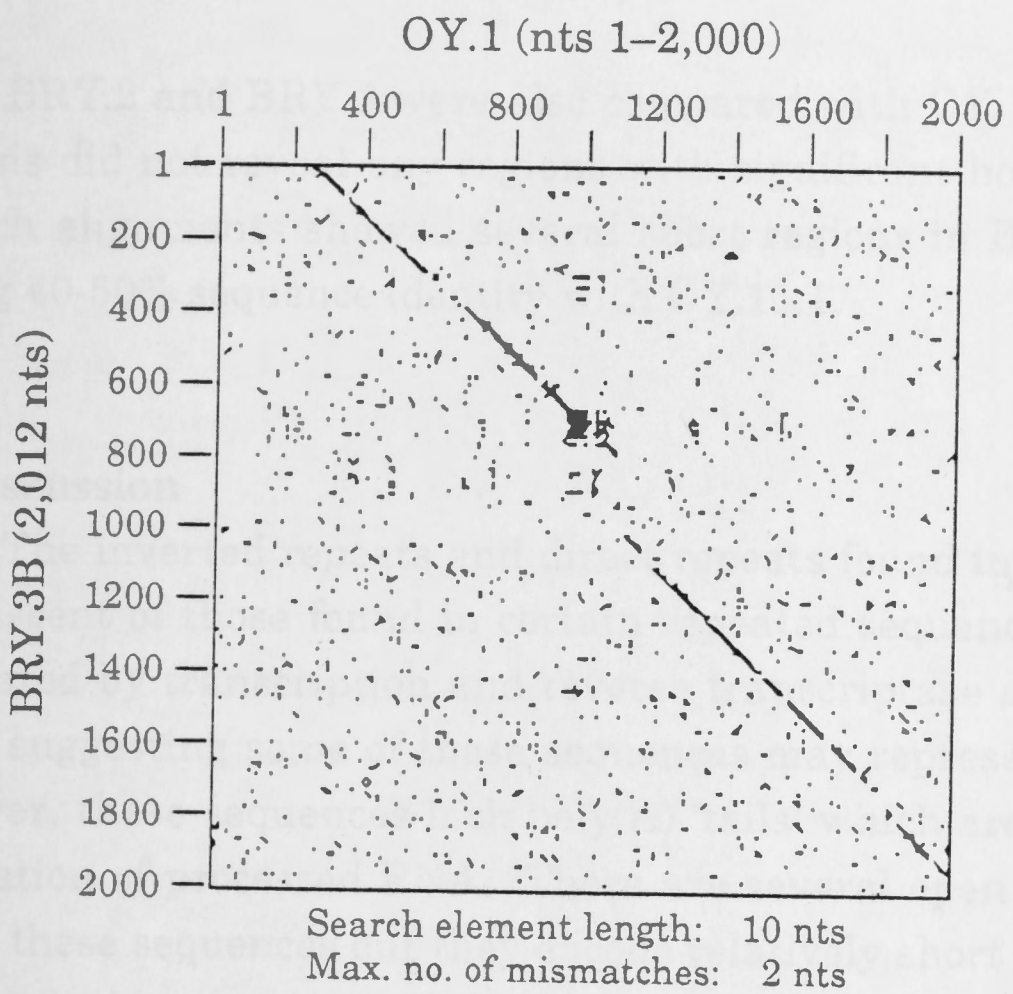


Figure 4.10. Dot matrix showing the region of homology between BRY.3B and OY1.

The homology between OY.1 and BRY.2 and BRY.3 is summarized in dot matrix form in Figure 4.11 for BRY.3B and BRY.2, which are contiguous in phage EMBL3A.Y1 (Chapter 3). The information for all of the comparisons with OY.1 is also summarized in Table 4.2. These summaries show that the cattle Y-chromosomal sequences contain inserts relative to OY.1, so that on the whole the nucleotide numbers for regions in OY.1 homologous to BRY.2 and BRY.3 are sequential, while the numbers for these regions in BRY.2 and BRY.3 show the presence of sequences which are absent from OY.1. The exception is for the region of homology in BRY.3A, where OY.1 contains a region which is absent from BRY.3A.

BRY.2 and BRY.3 were also compared with OY.11.1. Dot matrix analysis did not reveal any regions with significant homology. Needleman-Wunsch alignments showed several short regions in BRY.2 and BRY.3B having 40-50% sequence identity with OY.11.1.

4.4 Discussion

The inverted repeats and direct repeats found in these sequences are reminiscent of those found in certain repeated sequences which are duplicated by transcription and reverse transcriptase activity (Rogers, 1985), suggesting some of these sequences may represent pseudogenes. However, these sequences lack poly(A) 'tails' which are another sign of integration of processed RNA. There are several open reading frames within these sequences but they encode relatively short (about 100 amino acids) putative proteins and there does not seem to be a relationship between the possible polyadenylation sites and the open reading frames.

The sequences found in BRY.2 and BRY.3 also resemble the LINEs described in rodents (Brown and Dover, 1981) and primates. Some members of this family of sequences are 5.6 kb long in *Mus musculus* . The main primate family (the *Kpn*-LINE family) is 6.4 kb long and repeated three to five thousand times in the human genome (Adams *et al.*, 1980). Like the sequences in BRY.2 and BRY.3, the human and rodent repeats are generally interspersed among other sequences and do not appear to contain any major internal repetition. As with the primate *Kpn* family, the rodent sub-forms are distinguishable within each of several genomes by restriction site polymorphisms. Related genomes contain distinct divergent sub-

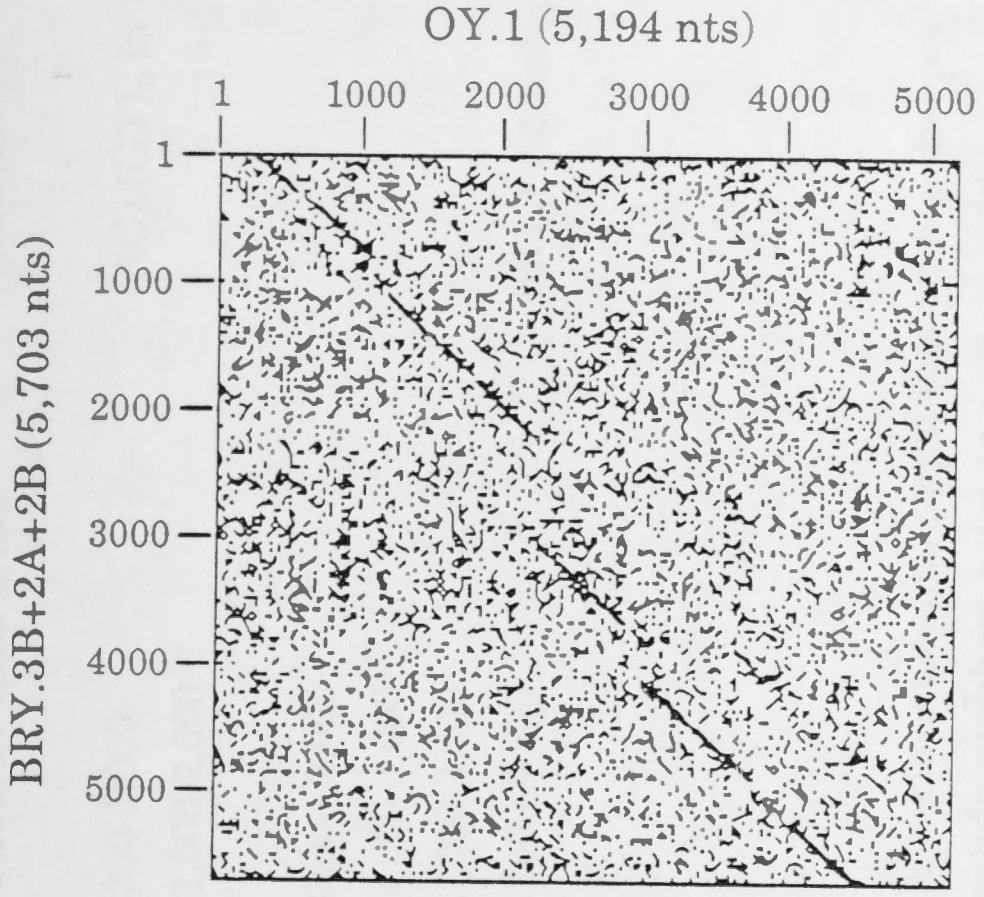


Figure 4.11. A summary of the regions in BRY.3B and BRY.2 which are contiguous in EMBL3A.Y1, showing the overall homology with OY1. As noted in the text the homologous regions are contiguous in OY1 but the cattle sequences have insertions relative to OY1.

Region of BRY.3/BRY.2	Region of OY1	Special features
BRY.3A: nt 804-1,360	nt 1-150, nt 4,150-5,194	
BRY.3B: nt 10-797	nt 300-1,000	$(AC).(GT)_n$ in both sequences
BRY.3B: nt 1,016-2,010	nt 1,000-2,000	
BRY.2: nt 1-200	nt 2,000-2,000	
BRY.2: nt 1,166-1,632	nt 2,060-3,000	
BRY.2: nt 2,074-2,665	nt 3,000-3,600	
BRY.2: nt 2,970-3,692	nt 3,600-4,500	BRY.1 homologue (both sequences)

Table 4.2. A summary of the comparisons of cattle Y-chromosomal sequences with a sheep Y-chromosomal homologue to BRY.2. The nucleotide numbers refer to regions within the cattle sequence and OY1 which have sequence identity. With the exception of the region of homology within BRY.3A the regions within OY1 are sequential, in contrast with those in BRY.3B and BRY.2, showing the cattle sequences contain inserted sequences relative to OY1.

families as well as varying relative amounts of sub-families (Singer, 1982a). The approximate total copy number of the rodent LINE family is similar from species to species. Most LINEs have variable 5' ends and common 3' ends, presumably reflecting their method of dispersion via reverse transcription beginning at 3' dA-rich sequences and subsequent retro-transposition of incomplete copies to new loci (Singer and Skowronski, 1985).

Homopurine sequences such as the $(GA)_{24}$,$(GAGGA)_6$,$(GA)_{21}$ tract seen in BRY.3, have been observed in rat genomes at the ends of the highly dispersed LINEs and have also been shown to arrest single stranded DNA replication in vitro (d'Ambrosio and Furano, 1987). It has been suggested that these homopurine sequences might play a role in transposition of the LINE elements by virtue of their ability to arrest DNA replication.

Although the sequences in the two subclones from this insert of Y-chromosomal DNA resemble LINEs in their length and apparent dispersion in the chromosome, it is difficult to envisage an RNA-mediated method of replication and dispersal which would confine them to one chromosome. However, Singer (1982b) proposed that the amplification and dispersal of LINEs may be via gene conversion, a non-reciprocal recombination in which a DNA sequence is duplicated at a (partly) distinct homologous site without being lost from the original 'donor' site. Singer further proposed that LINEs are maintained in smaller copy numbers than are SINEs (short interspersed repeats) because they are somehow limited to particular genomic regions. Both intra-chromosomal (Scherer and Davis, 1980) and inter-chromosomal (Jackson and Fink, 1981) gene conversion has been observed in yeast.

4.41 Dispersed repeats on the bovine Y chromosome

The non-Y-specific repeats within BRY.2 and BRY.3 probably include tracts of simple sequence. The presence of the sequences (TG).$(AC)_{11}$ in BRY.2 and (TG).$(AC)_{10}$ and (AC).$(TG)_{15}$ in BRY.3 may contribute to a pattern of hybridization indicating highly dispersed, high copy number repeats. One of the commonest naturally occurring repetitive sequences containing alternating purines and pyrimidines has C and A bases alternating on one strand and G and T on the other (Nordheim and Rich, 1983). Similar sequences have also been found in Y-chromosomal

sequences isolated from sheep genomic libraries using BRY.2, though this was not the only homology between the cattle and sheep sequences as revealed by sequence comparison.

Sequences of this type have been widely reported (Hamada *et al.*, 1982). A fragment containing the sequence $(CA)_n$ when used as a probe against an *Eco* RI digest of total mouse DNA gives a high level of 'background' hybridization (Nishioka and Leder, 1980) characteristic of a highly repeated sequence that occurs throughout the genome. Plaque hybridization studies of genomic libraries and a Genbank survey (Manor *et al.*, 1988) have indicated that these elements are spread throughout the genomes of primates and rodents and that they map primarily within introns and intergenic regions of these orders.

Long runs of $(CA)_n$ are also present in regions flanking known genes in the human genome (Nordheim and Rich, 1983; Kilpatrick *et al.*, 1984). There are about 50,000 copies of this sequence at least 50 bp long in human DNA and it is found in the same relative abundance in all eukaryotes examined (Hamada *et al.*, 1982). Conservation may be due to stringent selection. Nordheim and Rich (1983) demonstrated the formation of left-handed Z-DNA in sequences of $(dCA).(dGT)_{32}$ in negatively supercoiled plasmids. Z-DNA is a left-handed conformation of the double helix favoured by base sequences containing alternating purines and pyrimidines (Wang *et al.*, 1979), occurring in equilibrium with its more stable right-handed counterpart, the B-DNA form of duplex DNA. The B to Z transition occurs under physiological conditions (Nordheim and Rich, 1983). There seems to be one element of $(CA/GT)_n$ present per 50-100 kb. Since the DNA of eukaryotes is believed to be organized in domains in loops in which the DNA is attached periodically to a nuclear matrix, on average every 50-100 kb, there may be a functional correlation between copy number of $(CA/GT)_n$ segments and the number of DNA domains.

Slightom *et al.* (1980) suggested a $(CA/GT)_n$ sequence may be active in intergenic DNA exchanges in the human fetal γ-globin system and other examples suggest conformational flexibility of $(CA/GT)_n$ sequences could be instrumental in DNA rearrangements that are brought about by the specific action of recombinational proteins. These simple sequences appear to be 'hot spots' for recombination and gene conversion (Kilpatrick *et al.*,

1984). DNA is normally in a conformational equilibrium between right-handed B-DNA and left-handed Z-DNA. There is always a significant number of these sequences which exist in an untwisted version, neither left nor right-handed. These strands may be able to open up and separate and may recombine with other similar open strands. Z-DNA binding proteins have been identified in several organisms (e.g. Kmiec *et al.*, 1985; Blaho and Wells, 1987). These proteins promote homologous recombination but it is not known how this function relates to their Z-DNA binding capacity. Z-DNA may also contribute to gene regulation (Blaho and Wells, 1987). Z-DNA segments may be able to not only change the environment near a particular gene, but through interaction with supercoiling these segments could modify the transcribability of DNA regions far from the site at which the Z-DNA segment is found. The postulated role of Z-DNA in recombination is very interesting considering the presence of these sequences amongst the repeats of the cattle Y chromosome. It is also possible that conservation of $(CA)_n$ is due to similar pathways in all eukaryotes and does not necessarily imply this sequence has a selected function since $(CA)_n$ and other tandem repeats may be generated by a tandem amplification of existing sequences, perhaps by slippage during DNA replication.

The interspersion of sequences which are dispersed throughout the genome with repeats specific to the Y chromosome indicates that the cattle Y chromosome has not been genetically isolated for all of the period since it acquired the repeats which are now Y-specific. It is possible that the cattle Y chromosome may once have had a tandem repeat structure, but that during the much longer time since the original amplification events (in comparison with man) so much scrambling and interspersion of other sequences has occurred that the tandem blocks have been broken up. It has been suggested (Watson *et al.*, 1987) that tandem repeats of an identical DNA sequence would be an ideal substrate for intra-chromosomal recombination, which would lead to loss of tandemly repeated families.

4.42 Short-term organization of the cattle Y chromosome in comparison with the sheep Y chromosome

The homology between OY.1 and BRY.2 or BRY.3 is greater than that between BRY.2 and BRY.3. The sheep Y chromosome is smaller (12×10^6 bp, Lord, 1989) than the cattle Y chromosome (15×10^6 bp), so this result

may suggest that the original repeat unit , represented by OY.1, has been more vulnerable to scrambling and to growth in distance between the elements of the original Y-specific sequences on the cattle Y chromosome than it has been on the sheep Y chromosome. Elements which are contiguous in OY.1 are interspersed with other sequences in BRY.2 (summarized in Figure 4.11 and Table 4.2). OY.1.1, a 2,930 bp subclone from OY.1, is composed almost entirely of sequences which are only repeated on the Y chromosome; therefore the regions in BRY.2 and BRY.3 which have homology with this part of OY.1 can be related to the results of Chapter 3. OY.1.1 begins at nucleotide 2,264 in OY.1 and Table 4.2 shows that most of the second region of BRY.2 with homology to OY.1 has homology with OY.1.1, as do the third and fourth regions of BRY.2 having sequence identity with OY.1. The BRY.1 homologue is present in OY.1.1 at nucleotide 4,224, and the length of male-specific repeated sequences within BRY.2 is extended by this comparison with OY.1.1 to 722 bp. The subcloned fragment BRY.2A (Chapter 3), which appears to be repeated only on the Y chromosome, is present in the first region of homology with OY.1. Unfortunately it is not known if this region of OY.1 is repeated only on the Y chromosome.

OY.1.1 is present at 1200 copies in bulls compared to only 290 copies in rams (Lord, 1989), which may imply either that these sequences on the cattle Y chromosome have been amplified more often than their counterparts on the sheep Y chromosome or that they have been more fragmented and then amplified as part of a new unit, or both. BRY.2 is not present at higher copy numbers in sheep than it is in cattle (Sections 3.34 and 3.36). There is no clear relationship between the numerous direct repeats or the inverted repeats in the cattle sequences with the sequences which have been inserted relative to OY.1. OY.1 also contains an alternating purine and pyrimidine repeat $(AC).(GT)_{26}$ in the region which has homology to the region of BRY.3B which contains a similar stretch. Direct repeats and inverted repeats are also present within OY.1, perhaps reflecting the original means of migration to the progenitor Y chromosome.

The isolation of other repeated sequences specific to the bovine Y chromosome by other workers (Leonard *et al.*, 1987; Bondioli *et al.*, 1989) which are not Y-specific in sheep may reflect the fact that the cattle Y chromosome has undergone more changes than the sheep Y chromosome.

The extra size of the cattle Y chromosome may be due to the addition of more sequences, some of which have been amplified since to create new families of Y-enriched sequences.

The Y chromosome is not immune to colonization by sequences which spread by RNA intermediates, so the originally tandem Y-specific sequences may have become interspersed with the *Alu*-like artiodactyl repeats and perhaps the other non-Y-specific repeated sequences found within BRY.2 and BRY.3 are also sequences which do not rely on conventional methods of recombination for their dispersal.

The retroposition of non-viral cellular RNA species has emerged as a major evolutionary force, contributing to continuous sequence duplication, dispersion and rearrangement of eukaryotic genomes (Weiner *et al.*, 1986). In order to test this hypothesis other representative segments of DNA containing these Y-specific repeats should be isolated and analysed for evidence of uniform structure or a tandem repeat nature so that an indication of the original method of amplification on the Y (saltation giving a tandem repeat structure, or gene conversion via unequal intra-chromosomal recombination) can be obtained.

The repeated sequences within BRY.2 and BRY.3 which are also repeated elsewhere in the genome and which do not have sequence identity with OY.1 are not representatives of the artiodactyl SINE family, and they differ from one another so their origin and means of dispersal remains unknown. A search for evidence of transcription of the lower copy number non-Y-specific repeated sequences on the Y chromosome may also be useful in testing this hypothesis. Chapter 5 describes the isolation and characterization of several other recombinant phage containing sequences from the cattle Y chromosome.

CHAPTER FIVE

THE ISOLATION AND CHARACTERIZATION OF REPEATS COMPRISING 2% OF THE CATTLE Y CHROMOSOME

CHAPTER FIVE

The Isolation and Characterization of Repeats Comprising 2% of the Cattle Y Chromosome

5.1 Introduction

The repeated Y-chromosomal sequences found within EMBL3A.Y1 resemble the primate and rodent LINE families in several ways, being long repeats interspersed among other sequences with very little internal repetition (Bouchard, 1982). The copy number and length of the sequences also bear similarities to the LINEs. Homopurine-homopyrimidine tracts such as (CA).$(GT)_n$ and homopurine tracts are also seen in both mammalian LINE families and these bovine Y-chromosomal repeats. Another characteristic of the LINE families is that they show member sequence divergence. The primate *Kpn*-LINE family shows extensive sequence divergence among members, although the various members cross-hybridize. Both cloned human and monkey segments show restriction site variability (Singer, 1982). The isolation of other members of the family of bovine repeats specific to the Y chromosome may show whether this characteristic also applies to these bovine sequences.

The repeated sequences in BRY.2 and BRY.3 show some of the other features characteristic of RNA-derived inserted sequences, such as direct repeats. However they lack some features observed in some other LINE families, such as signs of RNA processing (Grimaldi *et al.*, 1984) and it is difficult to envisage how the Y-specific components of these sequences could have been confined to the Y chromosome over the long period observed if these sequences use the RNA-mediated methods of amplification used by other LINEs (Meunier-Rotival and Bernardi, 1984). It is possible that some of the Y-specific sequences were inserted on the Y chromosome after reverse transcription of an RNA and then underwent rapid duplication on the Y chromosome of a progenitor artiodactyl. It seems quite likely that the non-Y-specific repeats with which they are interspersed came to the Y chromosome later, via some method of transposition, as did the *Alu*-like artiodactyl short interspersed repeats (Watanabe *et al.*, 1982) surrounding BRY.2 and BRY.3 on the Y chromosome and found in other sequences from the bovine Y chromosome (K. Matthaei, personal communication).

The working hypothesis adopted is that the Y-specific repeats were translocated or retroposed from the X and/or autosomes in a ruminant progenitor. The newly acquired sequences then underwent amplification to produce tandem arrays which became interspersed by blocks of highly and moderately repeated sequences which are spread throughout the genome by an RNA-mediated process.

To test this hypothesis it was necessary to isolate other members of the family of repeats specific to the Y chromosome and look for evidence of a common structure, an underlying, perhaps long-period, tandem structure, common restriction sites, and variation among members. There was also a need to establish that the pattern of hybridization of these sequences to several *Bam* HI fragments of varying sizes does represent interspersion with other sequences on the Y chromosome and not simply the loss or addition of restriction sites. The isolation of recombinant phage containing overlapping sequences may provide information about tandem structure and the proximity of such blocks of sequences to one another. Chromosome walking was not possible because the Y-specific sequences are low to moderate copy number repeats and because there are blocks of highly repeated dispersed genomic sequences immediately adjacent to BRY.2 and BRY.3 on the Y chromosome.

Information about whether any of the repeated sequences contained within BRY.2-and BRY.3 are transcribed and could have contributed to the scrambling of a regular repeat structure by the insertion of a reverse transcribed copy of an RNA would also help to test the hypothesis. Several open reading frames of a reasonable length were found within BRY.2 and BRY.3. The subclone from EMBL3A.Y1 which hybridizes most strongly and specifically to male DNA, BRY.2, was used to re-probe the genomic cattle library.

5.2 Materials and Methods

5.21 Phage library screening, isolation and mapping of phage DNA

The EMBL3A genomic cattle phage library described in Chapter 3 was re-screened with BRY.2. 240,000 plaques were plated onto six large plates and nitrocellulose filter lifts made as described previously. The 3.7 kb *Bam* HI insert (BRY.2) was gel purified, and 0.6 μg was labelled by

nick-translation. An excess (3 mg) of unlabelled female genomic DNA, sheared in a French pressure cell to an average length of 250 bp, was added then the DNAs were denatured. Before addition to the hybridization mix (Chapter 3) the probe mixture was annealed at 65°C for 10 min. The use of unlabelled female DNA was to 'compete out' the non-Y-specific sequences in BRY.2 and prevent their hybridization to phage containing DNA from the X and/or autosomes. The filters from the primary screen were then stripped and re-probed with OY.1.1, a 2,930 bp subclone from the sheep homologue to BRY.2, OY.1 (Lord, 1989), which was also mixed with an excess of genomic cow DNA. One plaque hybridizing with this probe but not with BRY.2 was selected.

Phage were isolated as described in Chapter 3 and phage DNA was prepared as previously. The phage were mapped by end-labelling (Materials and Methods, Chapter 3) and by digestion with single restriction endonucleases and combinations of two endonucleases, followed by electrophoresis in agarose gels. As in Chapters 2 and 3 all digestions were carried out according to the instructions of the manufacturer.

5.22 Hybridization analysis of sequence elements

After digestion and electrophoresis for restriction site mapping the DNAs were transferred by alkaline blotting (Chapter 2) to Zeta-Probe membrane. The membranes were then probed with nick-translated BRY.1, BRY.2 and the male-specific fragment of BRY.2, BRY.2A. The filters were also probed with OY.1.1 or another subclone from a sheep Y-chromosomal DNA fragment isolated using BRY.2 as a probe from a ram genomic library, OY.11.1 (Lord, 1989). OY.11.1 is a subclone from the BRY.2 homologue which does not hybridize with OY.1 or BRY.2, but which is repeated specifically on the Y chromosomes of cattle, sheep and goats.

5.23 Hybridization with cDNA

The membranes carrying the restricted phage were also probed with labelled cDNA synthesized from poly(A^+) RNA from adult bull testis and with cDNA made from poly(A^+) RNA from foetal bull testis.

Poly(A^+) RNA was prepared by Sandra Beaton from RNA isolated from frozen adult or foetal bull testis (Cathala *et al.*, 1983) using a column of

oligo(dT)-cellulose (BRL) and a procedure modified from that described by Nakazato and Edmonds (1974).

Labelled cDNA was synthesized by denaturing 2.5 μg of RNA at 80°C for 90 sec with oligo(dT)$_{(12\text{-}18)}$ (Boehringer) at 0.02 mg/ml. The RNA was chilled and incubated in 50mM Tris-HCl, pH 8.3, 20mM DTT, 7.5mM Mg acetate, 100 μM each of dATP, dGTP and dTTP, 20 μM dCTP, 0.2 mg/ml BSA and 2 μl [α-^{32}P]dCTP (0.8 mCi/ml) with 1200 units/ml AMV reverse transcriptase (Life Sciences). The reaction was incubated at 42°C for 1 h.

Hybridization was at 70°C overnight in 2 x SSPE, 7% SDS, 0.5% BLOTTO (Johnson *et al.*,1984) and the filters were washed in 2 x SSC/0.1% SDS for 15 min at room temperature, followed by 30 min at 68°C in 0.2 x SSC/1% SDS.

5.3 Results

5.31 Phage containing 2% of the bovine Y chromosome isolated

Nineteen different recombinant phage containing inserts of cattle genomic DNA were isolated from a screen of 240,000 plaques from the EMBL3A phage library. Two of the phage were independently isolated twice, and one was isolated 5 times, showing the effects of amplification on the library. The phage were designated EMBL3A.Y2 through EMBL3A.Y20, following the nomenclature used for the original phage from which BRY.2 and BRY.3 were subcloned.

The phage DNAs were digested with *Bam* HI and *Hind* III, electrophoresed in an agarose gel, transferred to Zeta-Probe and hybridized with BRY.2 labelled by nick-translation. As expected (Figure 5.1), all the phage genomic inserts, except the one selected for lack of hybridization with BRY.2, hybridize with BRY.2, but since not all of the sequences within BRY.2 are repeated only on the Y chromosome, this fact alone does not prove the inserted DNA is from the Y chromosome. Most of the phage give a non-sex-specific dispersed pattern of hybridization when used to probe blots of digested genomic DNAs (data not shown), very similar to that obtained when the whole insert from EMBL3A.Y1 is used as a probe (Chapter 3, Figure 3.1). At least one of the phage (EMBL3A.Y8) did, however, show hybridization to male-specific bands in a *Bam* HI digest of

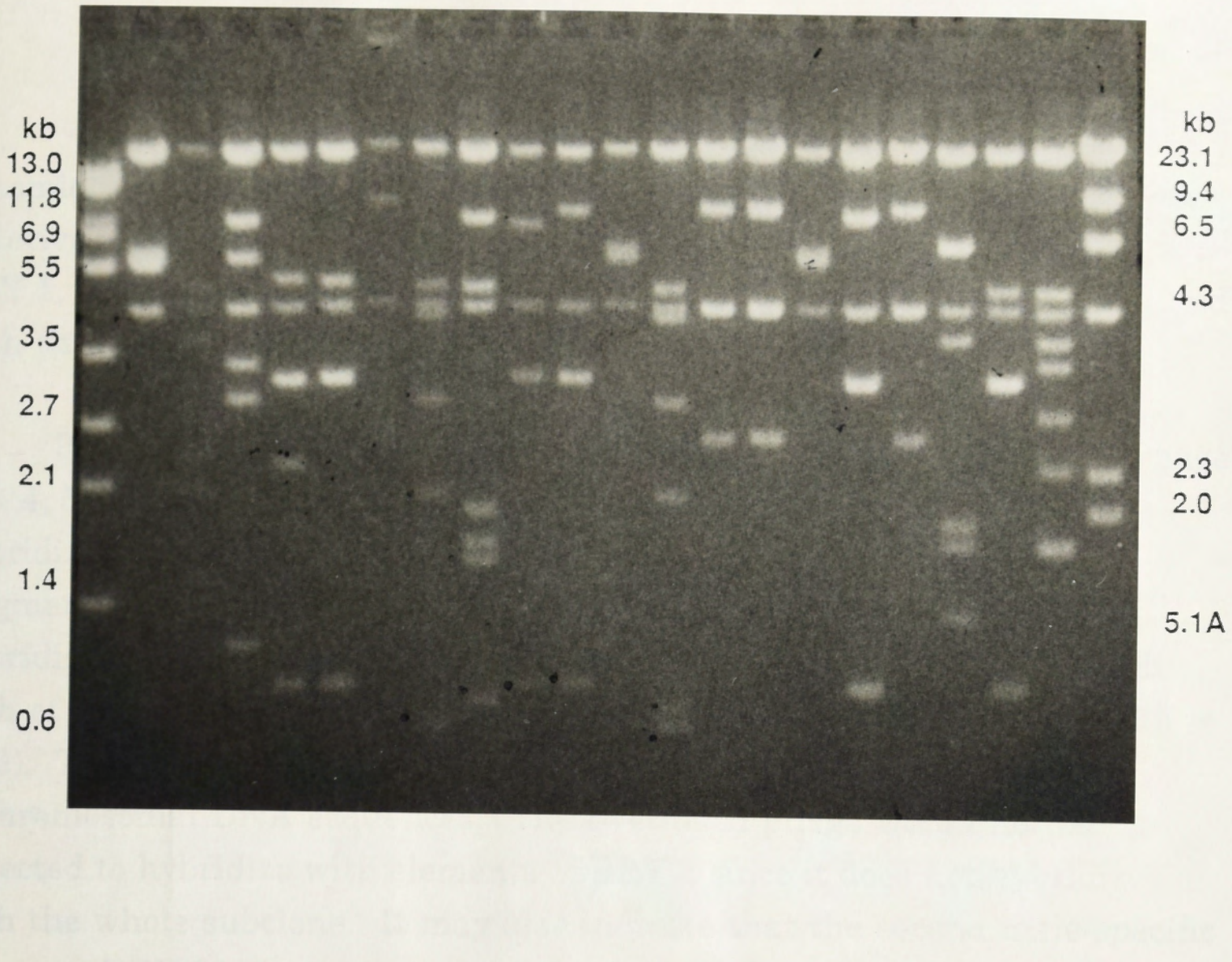


Figure 5.1 . The phage DNAs were digested with *Bam* HI and *Hind* III, electrophoresed in a1% agarose gel (**5.1A**), with λ *Acc* I and λ *Hin*d III markers (lanes 1 and 22), then transferred to Zeta-Probe. The filter was hybridized with nick-translated BRY.2 (**5.1B**). All of the phage except EMBL3A.Y20 hybridize with the probe.

genomic DNA (Figure 5.2). In order to establish whether the inserts were Y-derived, blots of the digested, electrophoresed phage were hybridized with BRY.1, which is mainly associated with the Y chromosome (Chapter 2) and with the male-specific fragment BRY.2A, subcloned from BRY.2.

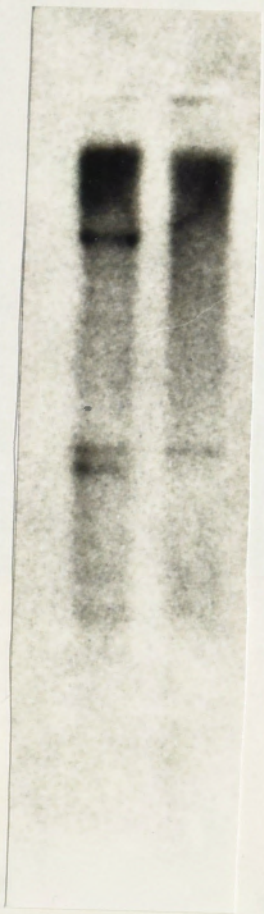
Phage 6, 8, 9, 12, 14 and 16 hybridized with BRY.1 (Figure 5.3). Phage 2, 3, 4, 5, 7, 10, 11, 12 (very faintly), 13, 15, 16 (very faintly), 17, 18 and 19 hybridized with BRY.2A (figure 5.4). Phage 15 and 18 contain sub-fragments which hybridize strongly with BRY.2 but which show no hybridization with BRY.2A or with BRY.1. The sub-fragments to which each of the probes hybridized are shown on the phage maps (Figures 5.10 to 5.20). This shows that all of the nineteen phage isolated contain cattle Y-chromosomal DNA sequences. The twentieth phage would not be expected to hybridize with elements of BRY.2 since it does not hybridize with the whole subclone. It may also indicate that the second male-specific sequence BRY.2A, is not represented on the Y chromosome in equivalent numbers with BRY.1. Only two of the phage, numbers 12 and 16, hybridized with both BRY.1 and BRY.2A and these two hybridize only very faintly with BRY.2A. In BRY.2 these two male-specific sequences BRY.1 and BRY.2A are at opposite ends of the 3.7 kb subclone, and this result seems to confirm that these sequences are almost always separated by other DNA sequences. In EMBL3A.Y12 the two Y-specific repeated sequences map to almost adjacent restriction fragments (Figure 5.16), while in EMBL3A.Y16 they map to one large *Bam* HI/*Hind* III fragment (Figure 5.18). If BRY.2A is present in lower copy numbers than BRY.1, it would further indicate that these groups of Y-associated sequences are heterogeneous.

The phage have an average insert length of 15 kb, which means if all have originated from the Y chromosome, and including the original phage, EMBL3A.Y1, that clones representing 15 kb x 20 =300 kb or $100/(1.5 \times 10^7$ bp/3×10^5 bp) = ~2% of the cattle Y chromosome has been cloned. The phage isolated because of its hybridization with OY.1.1 has been included in this calculation because OY.1.1 is male-associated in cattle, and the insert has been assumed to originate from the Y chromosome.

The phage were probed with the sheep Y-chromosomal sequence OY.1.1, isolated by homology with BRY.2. OY.1.1 gives a predominantly

1 2

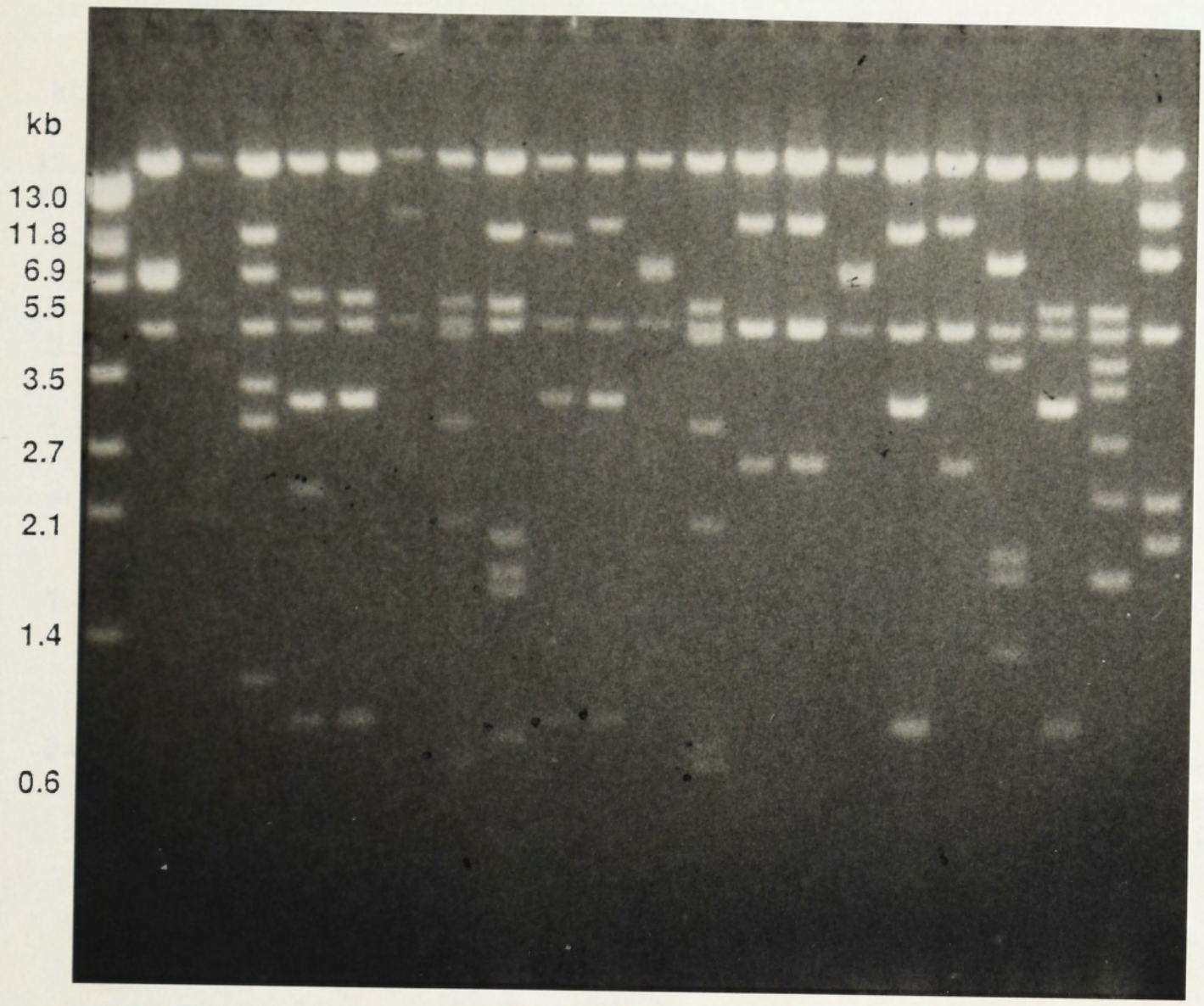
5.2A



1 2

5.2B

Figure 5.2 . Genomic cattle male (1) and female (2) DNA was digested with *Bam* HI, electrophoresed in a 0.8% agarose gel (**5.2A**), then transferred to Zeta-Probe. The filter was hybridized with EMBL3A..Y8 labelled by nick-translation (**5.2B**).



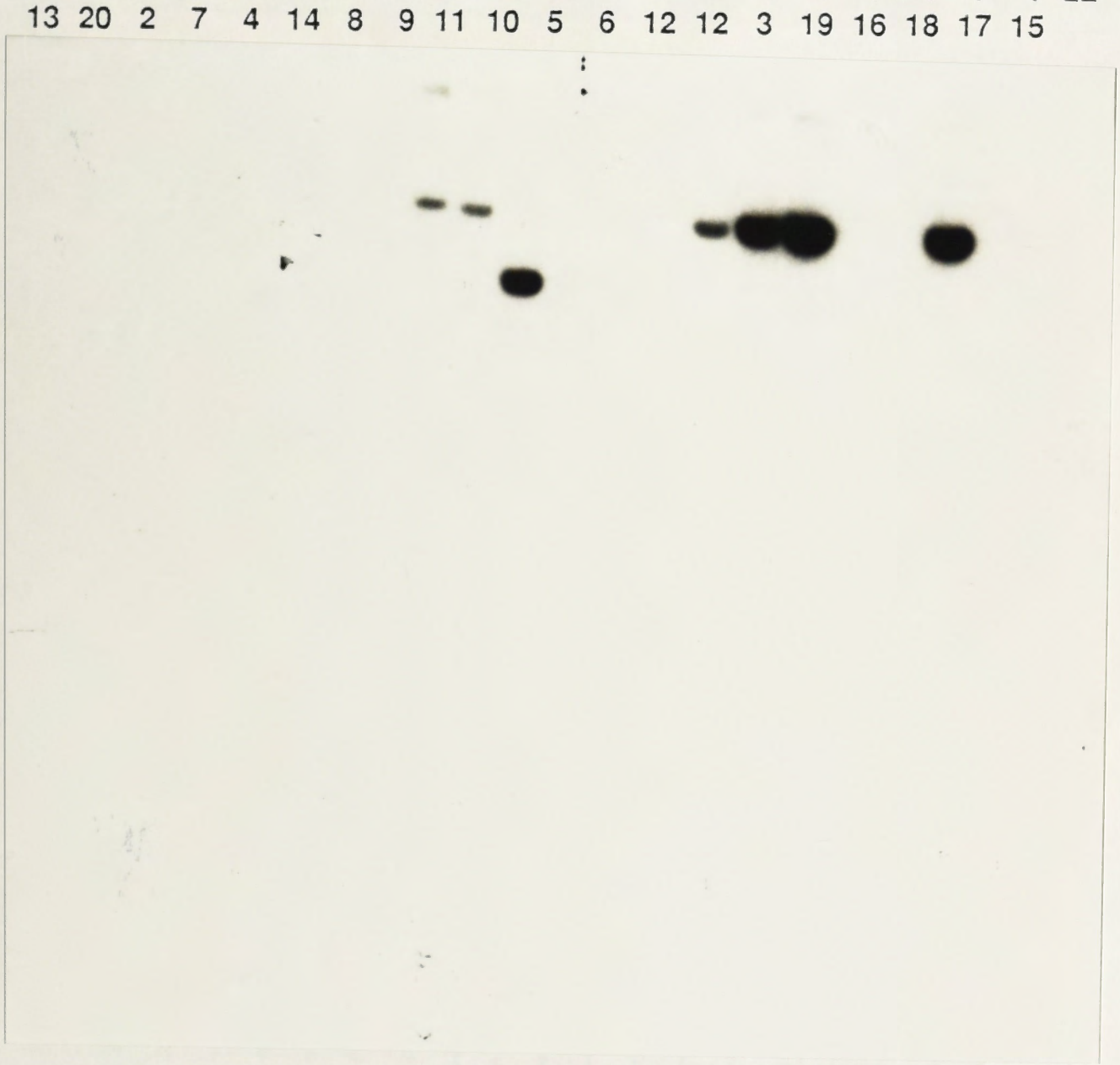


Figure 5.3 . The phage DNAs were digested with *Bam* HI and *Hind* III, electrophoresed in a1% agarose gel (**5.3A**), with λ *Acc* I and λ *Hin*d III markers (lanes 1 and 22), then transferred to Zeta-Probe. The filter was hybridized with nick-translated BRY.1 (**5.3B**).

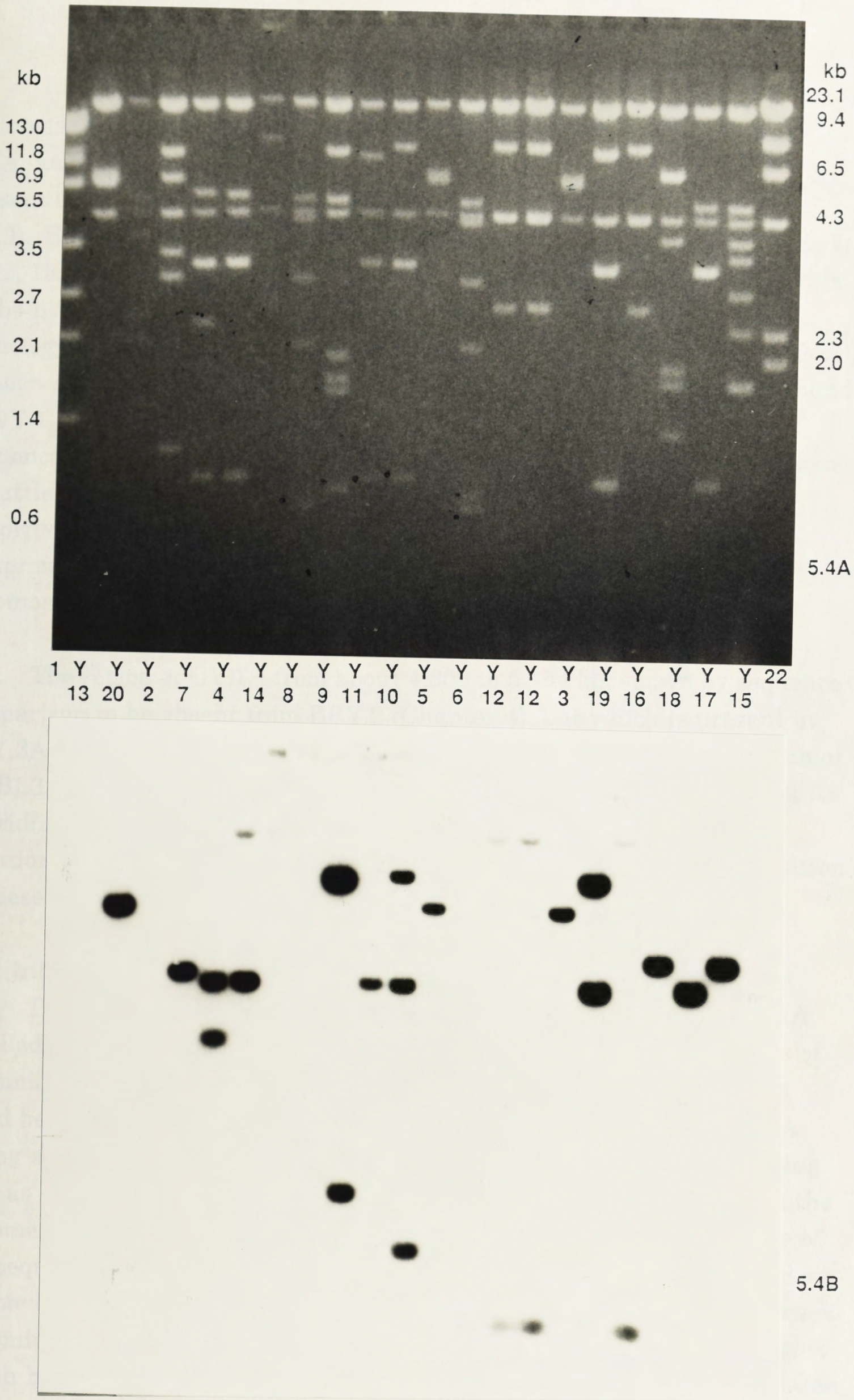


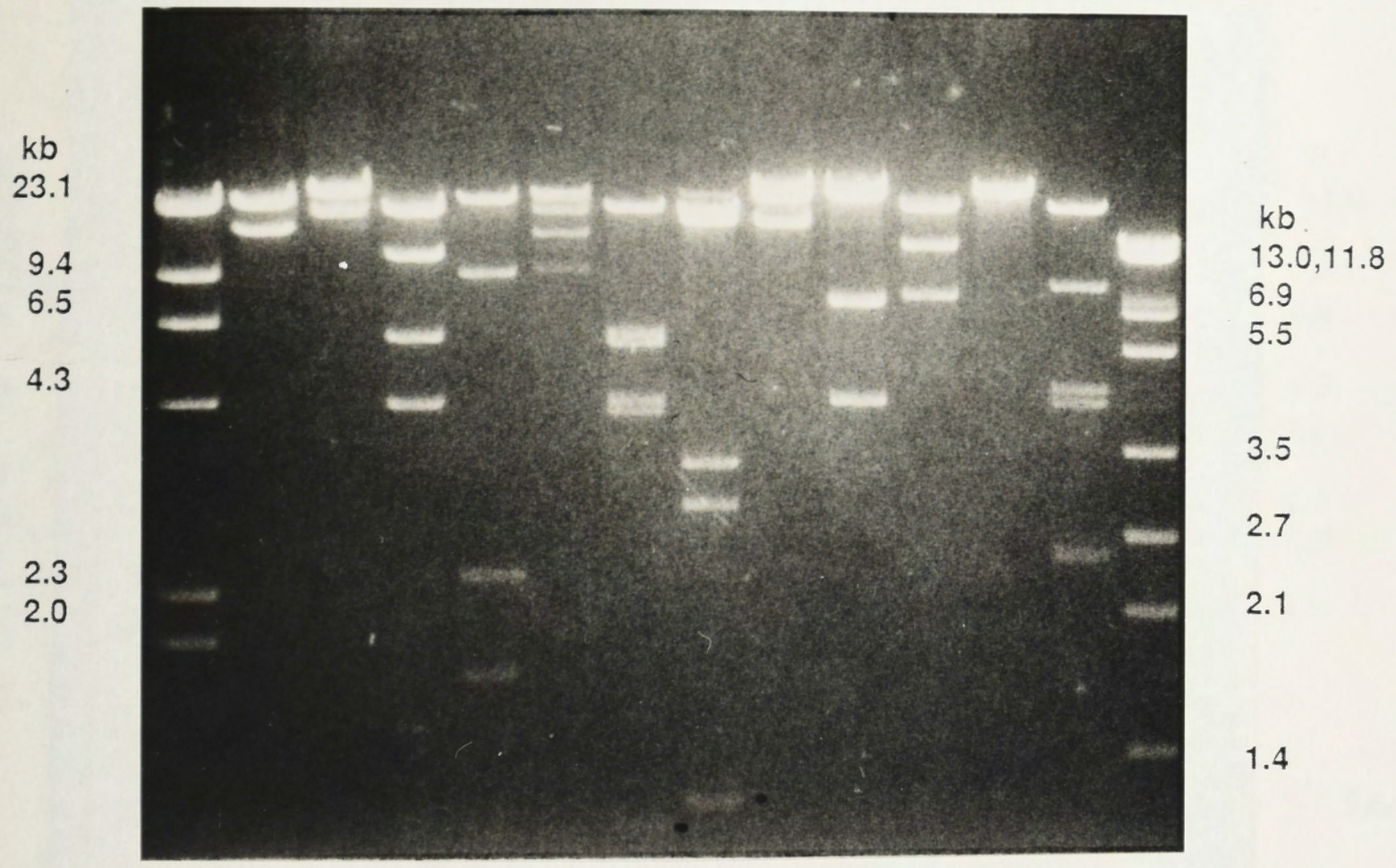
Figure 5.4. The phage DNAs were digested with *Bam* HI and *Hind* III, electrophoresed in a1% agarose gel (**5.4A**), with λ *Acc* I and λ *Hin*d III markers (lanes 1 and 22), then transferred to Zeta-Probe. The filter was hybridized with nick-translated BRY.2A (**5.4B**).

male-associated pattern when hybridized with genomic digests of cattle DNA, and in fact it is more highly repeated in cattle than in sheep, being present at about 1200 copies in the bull and 290 copies in the ram (Lord, 1989). Some of the data is shown in Figure 5.5. Phage numbers 2, 4, 5, 6, 7, 8, 12, 13, 14, and of course 20, hybridized with this probe. The fact that this probe hybridizes to restriction fragments of the phage not showing homology with either BRY.2A (phage 2, 5, 7, and 14, 20) or BRY.1 (6 and 20), as seen on the phage maps, indicates that some of the other elements found in OY.1.1 and BRY.2 are shared, as shown by comparison of their sequences (Chapter 4). Since OY.1.1 is repeated only on the Y chromosome in cattle these other elements must also be male-specific repeats. The majority of the phage contain an assortment of the sequences which comprise BRY.2 and OY.1.1 which have been conserved on the Y chromosome between the two species.

The region of OY.1.1 from about 4,800 to 5,194 bp, shown by sequence comparison to be absent from BRY.2 (Chapter 4), but which is present in BRY.3A (Chapter 4), is probably the element responsible for the isolation of EMBL3A.Y20 which was selected from the phage library on the basis of its hybridization with OY.1.1 and total lack of homology with BRY.2. The isolation of this phage is further evidence of the heterogeneous composition of these *Bam* HI fragments from the cattle Y chromosome.

5.32 Interspersed genomic repeats are common on the Y chromosome

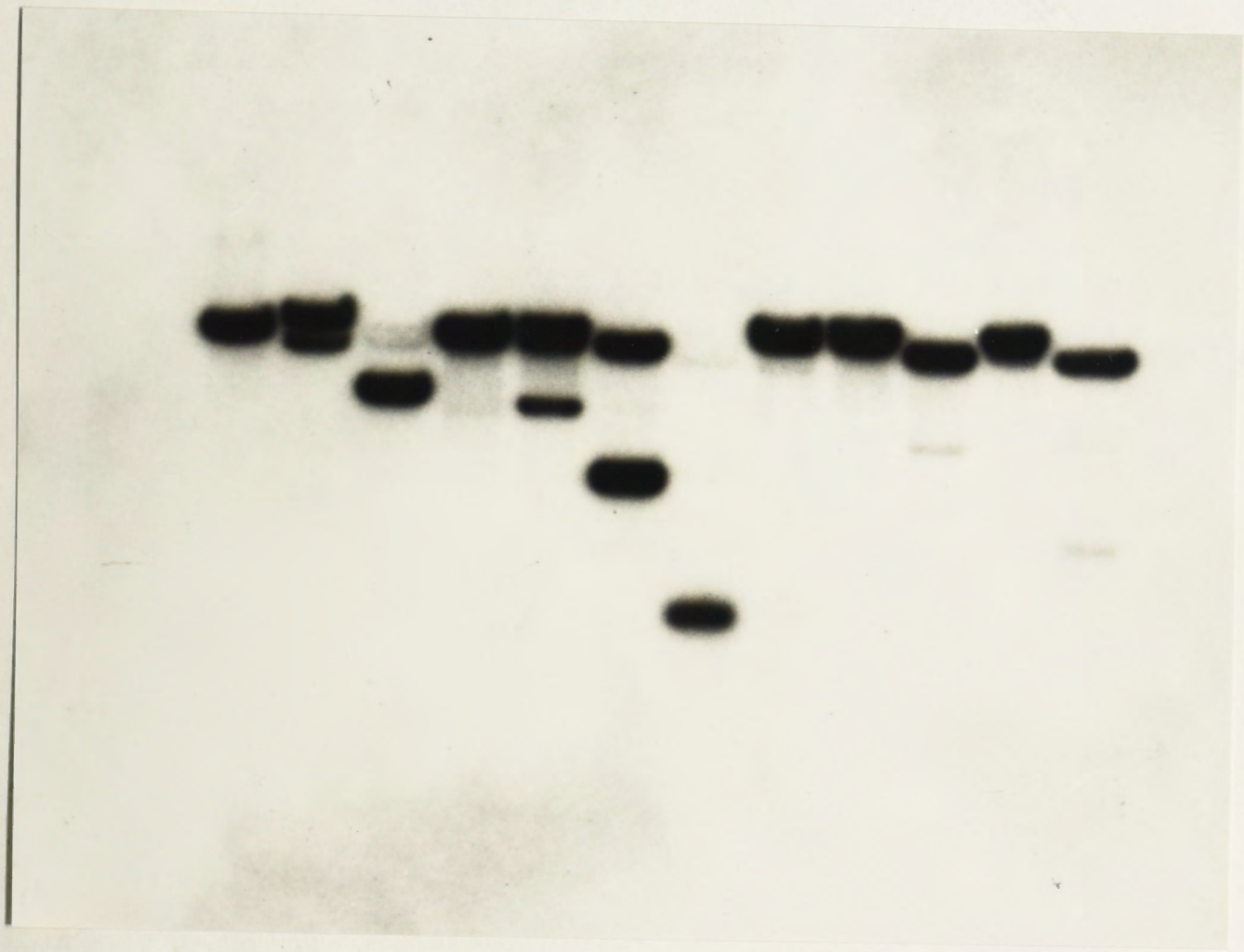
The blots of digested phage DNAs were also probed with cow DNA labelled by nick-translation (Figure 5.6), so that an idea of the numbers of genomic repeats found interspersed among these Y-specific sequences could be obtained. Twelve of the phage hybridized with the female DNA giving a strong signal after a short period of exposure to film, confirming that as seen in EMBL3AY.1, the presence of repeats from elsewhere in the genome interspersed with the Y-specific sequences is a general feature of the sequence organization. This result was expected since few of the phage are obviously from the Y chromosome when used to probe Southern blots of genomic DNA. This was confirmed by mapping the restriction fragments which hybridize with the genomic female DNA and comparing the position of these fragments with the fragments to which male-specific probes hybridize (Figures 5.10 to 5.20 below). Phage 1, 2, 3, 5, 6, 8, 9, 12, 13, 15, 16



5.5A

1 2 3 4 5 6 7 8 9 10 11 12 13 14

Y13 Y6 Y2 Y12



5.5B

Figure 5.5. Phage DNAs were digested with *Bam* HI, (lanes 2, 5, 8 and 11), *Eco* RI (lanes3, 6, 9 and 12), or *Hind* III (lanes 4, 7, 10 and 13), and electrophoresed in a 1% agarose gel with λ *Acc* I (lane 14) markers (**5.5A**). The DNAs were transferred to Zeta-Probe and hybridized with OY1.1, labelled by nick-translation (**5.5B**).

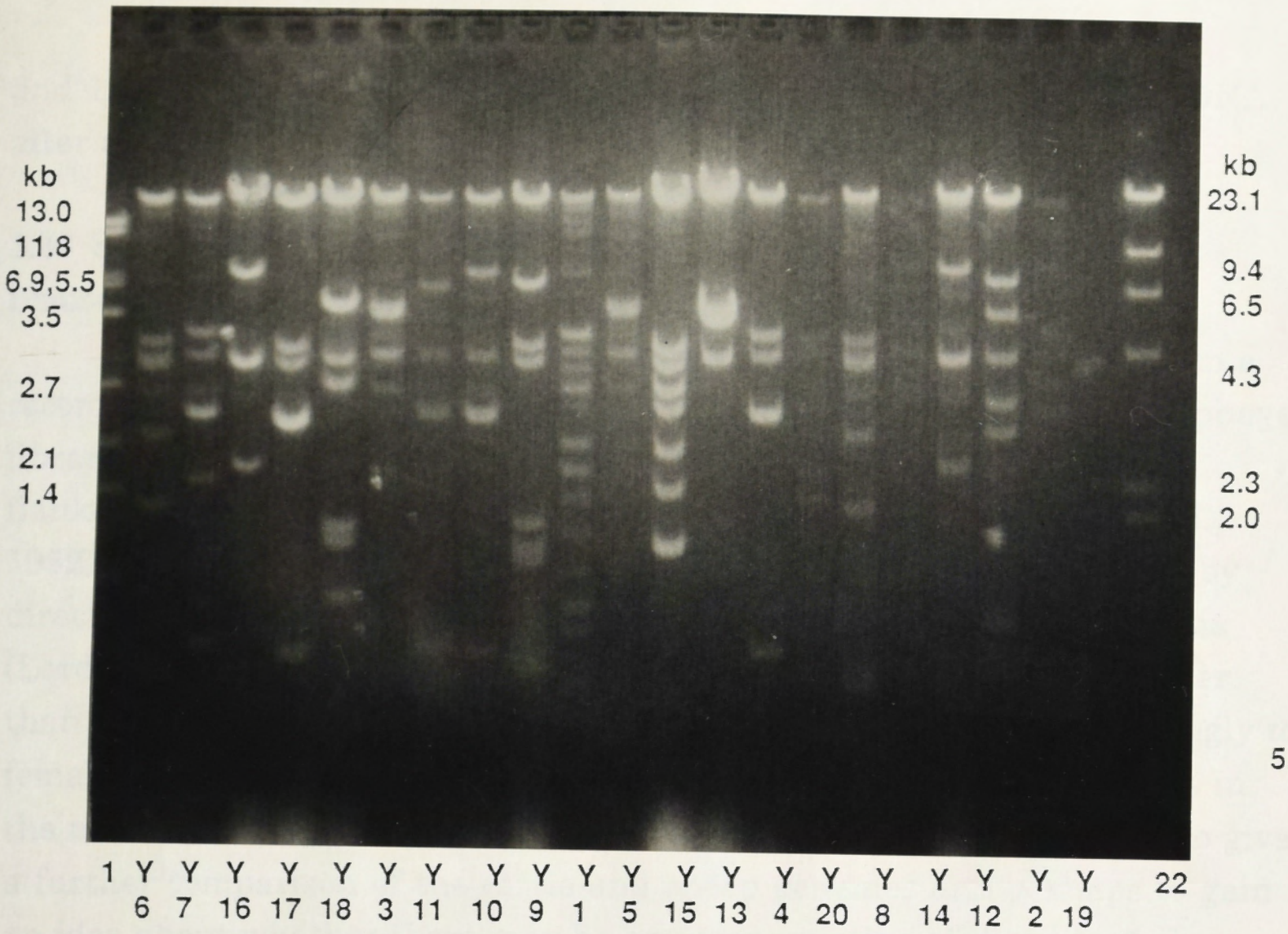


Figure 5.6. The phage DNAs were digested with *Bam* HI and *Hind* III and electrophoresed in a 1% agarose gel with λ *Acc* I and λ *Hind* III markers (lanes 1 and 22), (**5.6A**). The DNAs were transferred to Zeta-Probe and probed with female genomic DNA labelled by nick-translation (**5.6B**).

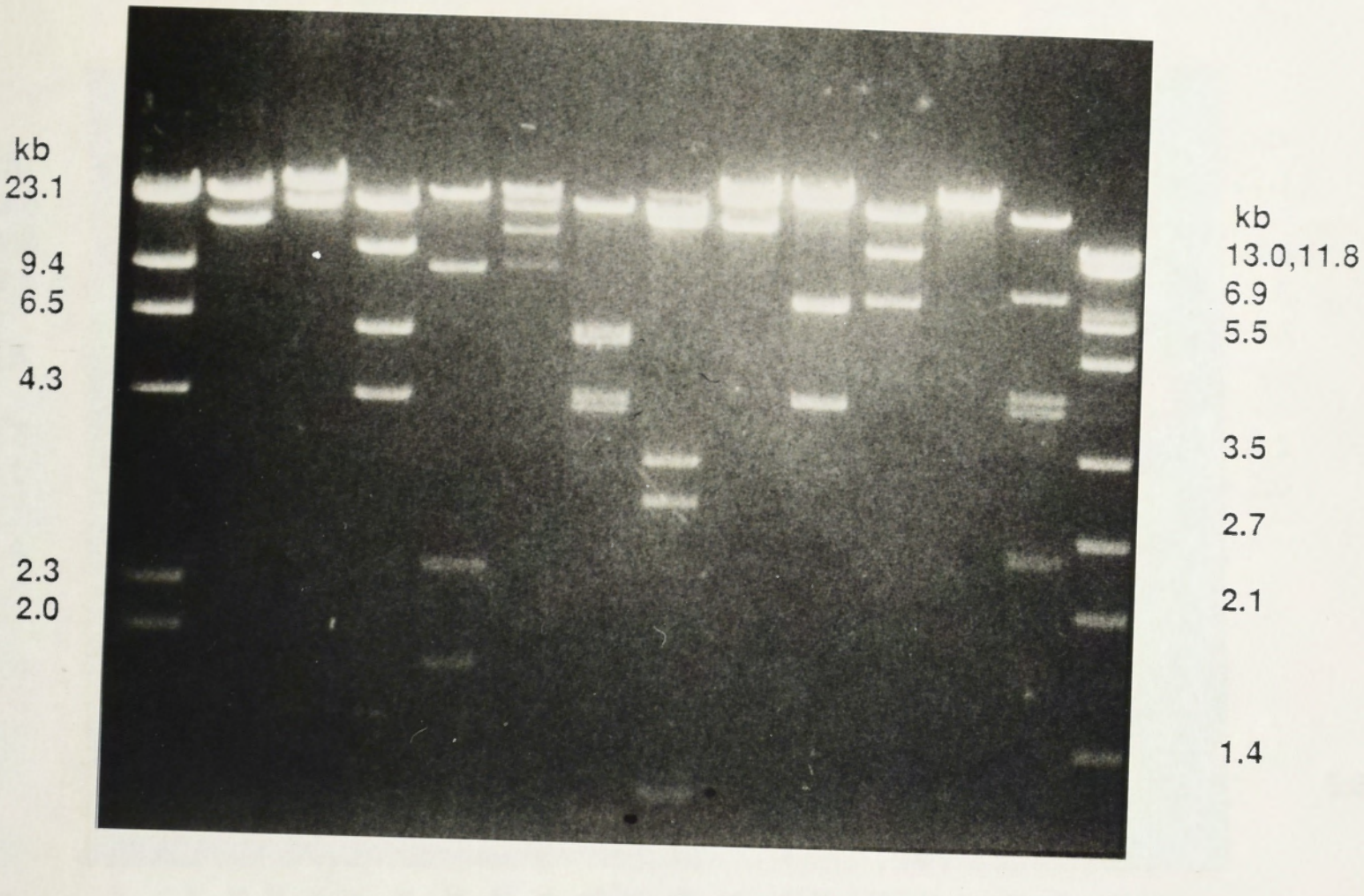
and 18, show hybridization with the nick-translated female genomic DNA after an overnight exposure to film at room temperature.

5.33 Sequences similar to repeated Y-chromosomal sequences are transcribed

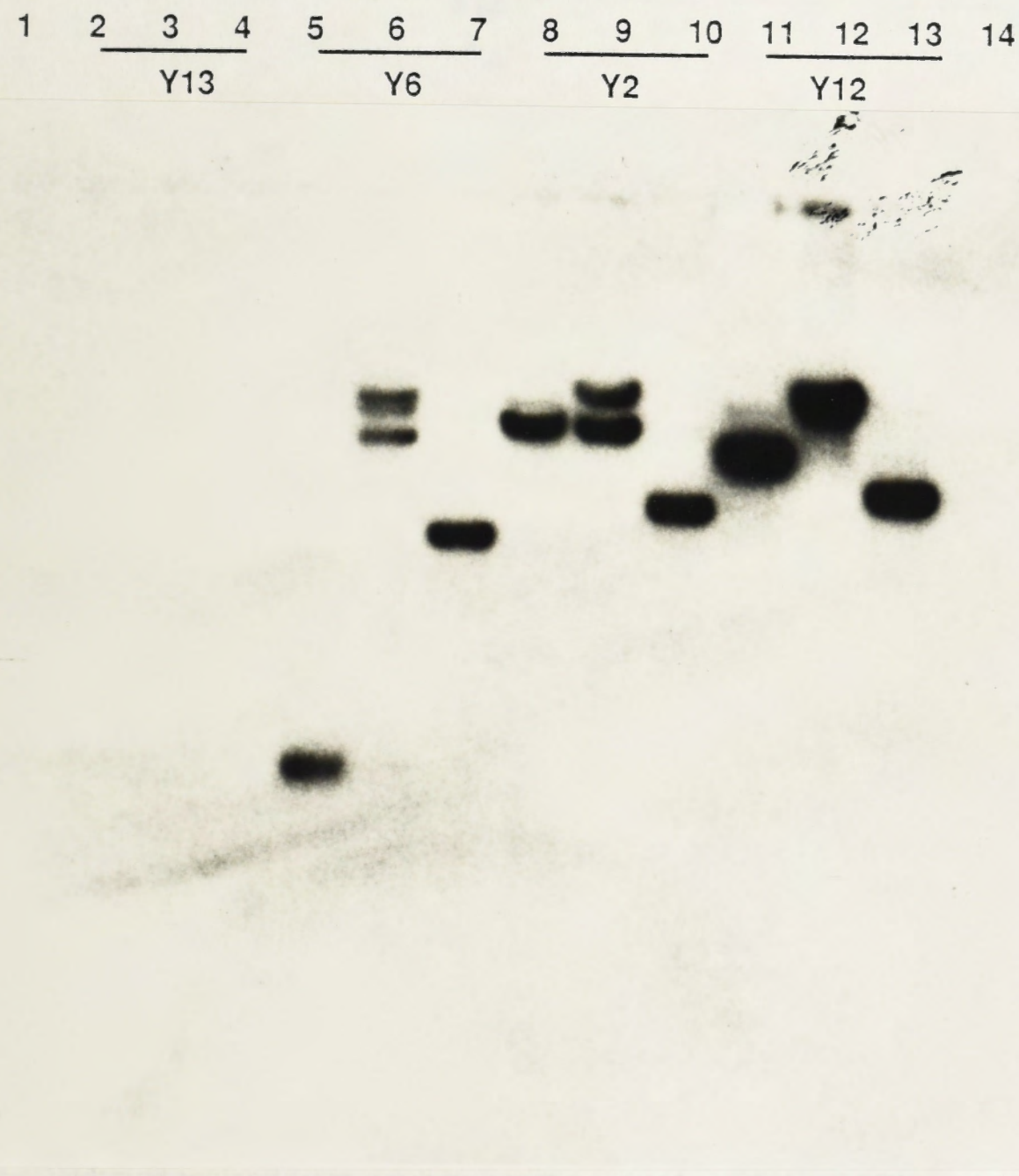
The phage were also probed with OY.11.1, another subclone from a recombinant phage isolated by Eric Lord from a genomic sheep DNA phage library using BRY.2. This sequence contains an open reading frame flanked by direct repeats and may represent a pseudogene (Hollis *et al.*, 1982) since it hybridizes faintly with foetal bull testis cDNA, is flanked by direct repeats and has homology with a transcript from adult bull testis (Lord, 1989). This sequence hybridizes predominantly with male rather than female DNA in sheep and cattle, although it hybridizes less strongly to female sheep DNA than cow DNA, and is present at about 1,000 copies in the male of both species. The purpose of this experiment was simply to give a further comparison of the cattle and sheep genomes and perhaps to gain an idea about whether there may be any transcription of the repeated sequences on the cattle Y chromosome.

As expected from sequence comparisons (Chapter 4), OY.11.1 does not hybridize with EMBL3A.Y1. Several of the phage (2, 6 and 12) did hybridize with OY.11.1 (Figure 5.7). In all of these phage restriction fragments hybridizing with OY.11.1 also hybridized with OY.1.1. In phage EMBL3A.Y2 (Figure 5.11) and Y12 (Figure 5.16) the restriction fragments to which OY.11.1 hybridizes are at the opposite end of the 19.2 kb and 15 kb inserts from the fragments which hybridize with the male-specific sequences BRY.1 and BRY.2A, though there may be a small overlap. There is definitely no overlap in EMBL3A.Y6 (Figure 5.13).

This result indicates homology of sequences on the cattle Y chromosome with at least one sequence in the sheep genome which is transcribed. The phage, including EMBL3A.Y1, from which BRY.2 and BRY.3 were derived, were probed with labelled cDNA synthesized from adult bull testis poly(A^+) RNA, Figure 5.8. Many of the phage (1, 2, 3, 4, 5, 6, 8, 9, 10, 12, 13, 14, 15, 16, 17, 18 and 19 very faintly) hybridized with the adult bull testis cDNA.



5.7A



5.7B

Figure 5.7. Phage DNAs were digested with *Bam* HI, (lanes 2, 5, 8 and 11), *Eco* RI (lanes 3, 6, 9 and 12), or *Hind* III (lanes 4, 7, 10 and 13), and electrophoresed in a 1% agarose gel with λ *Hind* III (lane 1) and λ *Acc* I lane 14) markers (**5.7A**). The DNAs were transferred to Zeta-Probe and hybridized with OY11.1, labelled by nick-translation (**5.7B**).

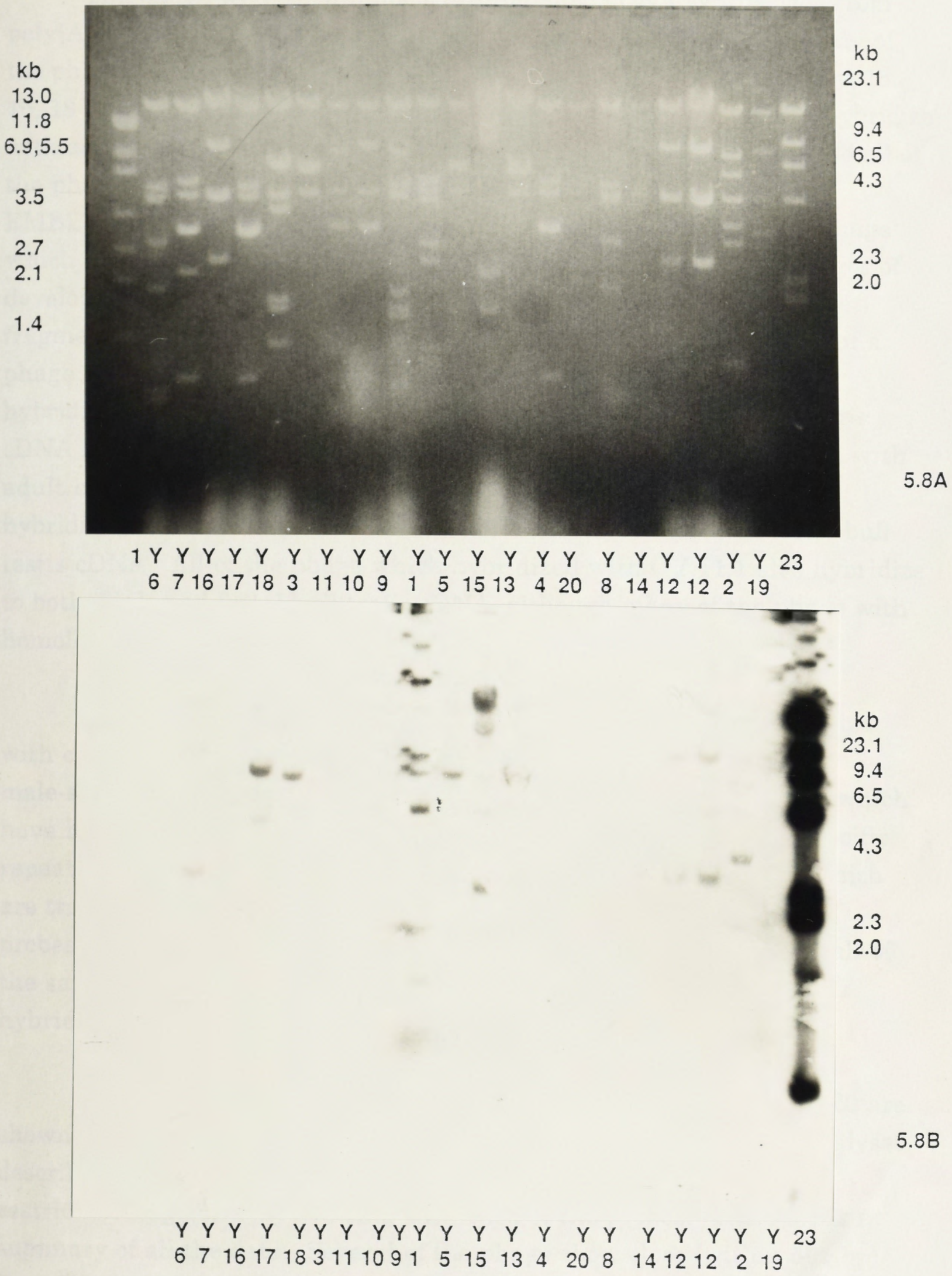


Figure 5.8. The phage DNAs were digested with *Bam* HI and *Hind* III and electrophoresed in a 1% agarose gel with λ *Acc* I and endlabelled λ *Hind* III markers (lanes 1 and 23), (**5.8A**), then transferred to Zeta-Probe. The filter was probed with labelled cDNA synthesized from adult bull testis poly (A+) RNA, (**5.8B**).

The phage were then probed with labelled cDNA synthesized from poly(A+) RNA isolated from foetal bull testis (Figure 5.9). Again many of the phage hybridized to the cDNA (1, 2, 3, 4, 5, 6, 8, 9, 12, 13, 14 faintly, 15, 16, 18 and 20). It should be noted that several phage (10, 17 and 19) hybridize with adult bull testis cDNA but not with foetal bull testis cDNA, while all of the phage which hybridize with foetal bull testis cDNA, except EMBL3A.Y20, also hybridize with adult bull testis cDNA. Several phage which hybridize with cDNA made from poly(A+) RNA from both stages of development do however show hybridization of different restriction fragments to message from each stage. EMBL3A.Y3 is an example of a phage containing an insert of DNA which shows such stage-specific hybridization. In the insert of this phage (Fig 5.11) the foetal bull testis cDNA hybridizes to two restriction fragments which do not hybridize with adult bull testis, but are immediately adjacent to fragments which hybridize with adult bull testis cDNA but do not hybridize with foetal bull testis cDNA. All of the phage which hybridized with OY.11.1 also hybridize to both adult and foetal bull testis cDNA, although many of the phage with homology to the cDNA probes have no homology with OY.11.1.

It should be noted that the restriction fragments which hybridize with cDNA are generally not the fragments which hybridize to the male-specific repeat probes (with the exception of those three phage which have homology with OY.11.1), suggesting that mainly the non-Y-specific repeated sequences have homology with sequences in the genome which are transcribed. This is further reinforced by the fact that the cDNA probes, in particular the foetal bull testis cDNA, hybridize to almost all of the same restriction fragments to which the labelled female DNA hybridizes.

The phage maps of phage EMBL3A.Y1 through to EMBL3A.Y20 are shown in Figures 5.10 to 5.20. The results of all the hybridization analyses described above are included on the maps as a line indicating the restriction fragments hybridizing with each of the probes. Table 5.1 is a summary of all the data. Several of the phage show close (within 500 bp) *Bam* HI and *Eco* RI sites, 1, 4 and, 6, 7, 9, 11,17, 18 and 19 and 20, while several have close *Hind* III and *Eco* RI sites: 1, 4, 5, 6, 7, 10, 15, and 20. Phage numbers 1, 4, 6, 12 and 20 have close *Bam* HI and *Hind* III sites. In phage numbers 6, 7 and 20 the three sites are clustered, however the

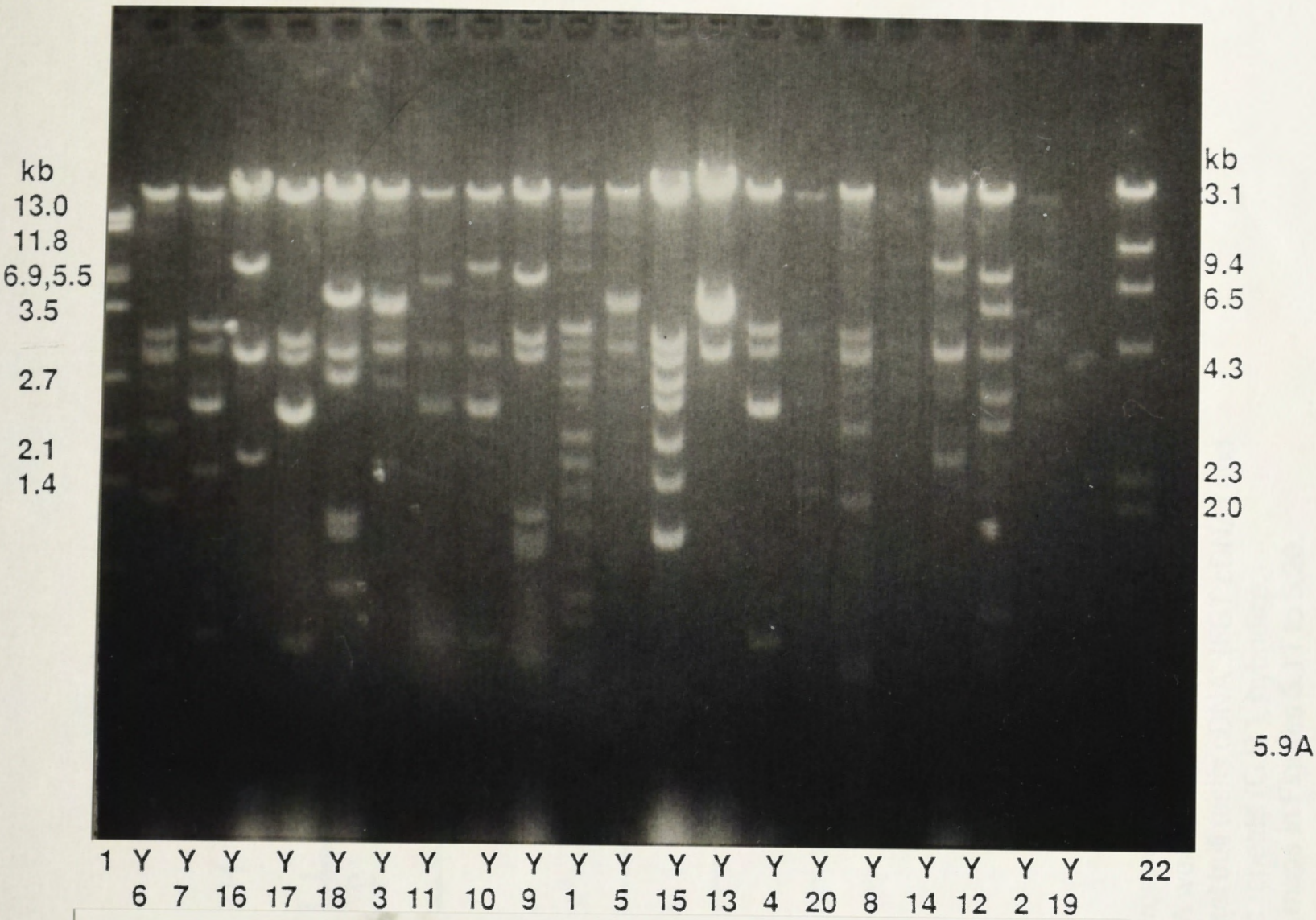
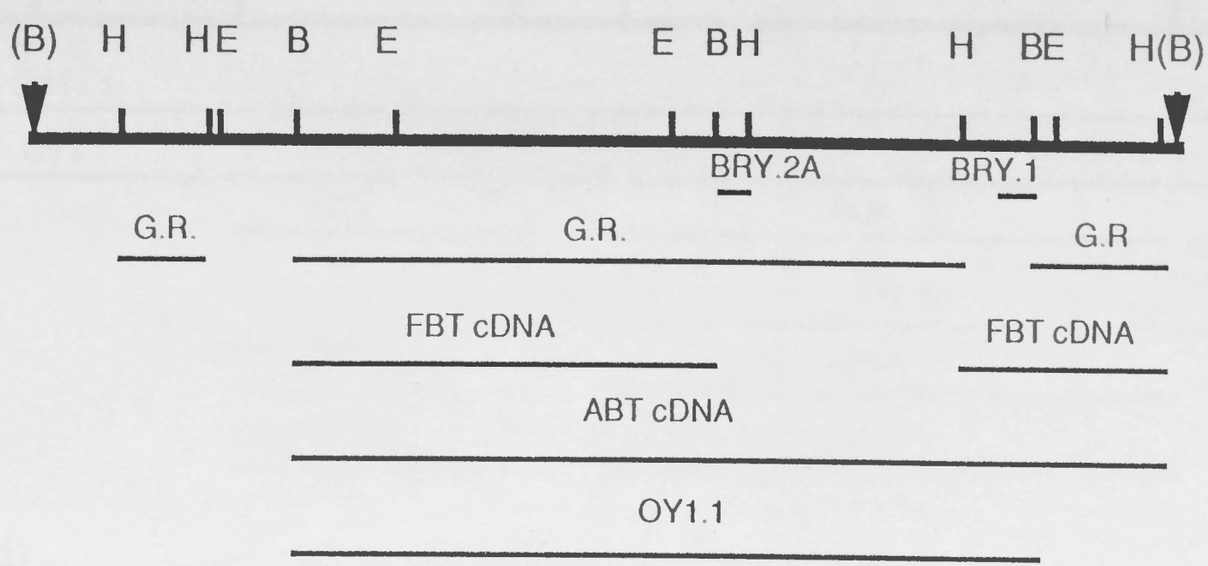


Figure 5.9. The phage DNAs were restricted with *Bam* HI and *Hind* III and electrophoresed in a 1% agarose gel with λ *Acc* I and endlabelled λ *Hind* III markers (lanes 1 and 22), (**5.9A**), then transferred to Zeta-Probe. The filter was probed with labelled cDNA synthesized from foetal bull testis poly (A+) RNA, (**5.9B**).



(B)- *Bam* HI site not regenerated by cloning
B - *Bam* HI site
H - *Hind* III site
E - *Eco* RI site

1 cm = 1 kb

Figure 5.10. Map of the insert of EMBL3A.Y1 (from which BRY.2 and BRY.3 were subcloned), showing the regions in which BRY.1 and BRY.2A were found by hybridization and sequencing and the regions to which OY1.1, foetal bull testis cDNA (FBT cDNA) and adult bull testis cDNA (ABT cDNA), and genomic repeats (G.R.), hybridize.
The scale and abbreviations apply also to the maps in Figures 5.11 to 5.20.

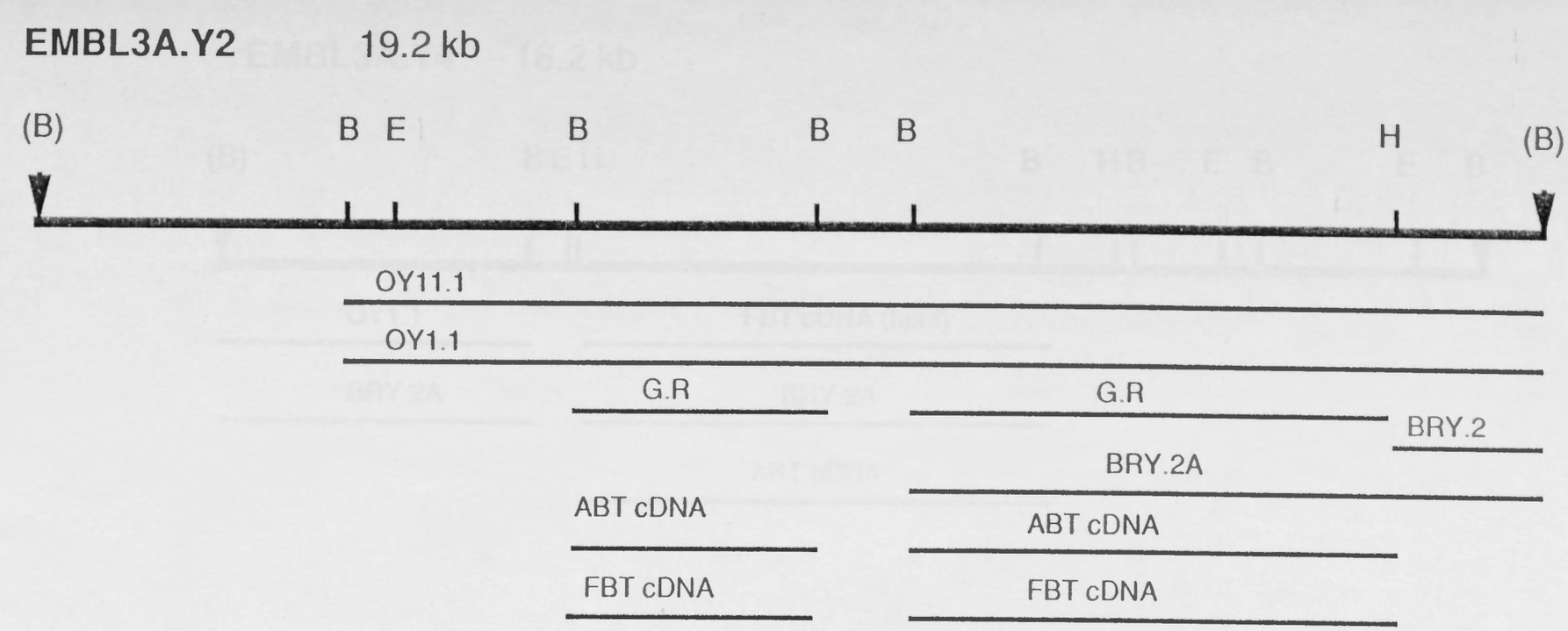


Figure 5.11. Restriction maps of EMBL3A.Y2 and EMBL3A.Y3, showing the fragments to which the various probes hybridize. The scale and abbreviations as are for Figure 4.10.

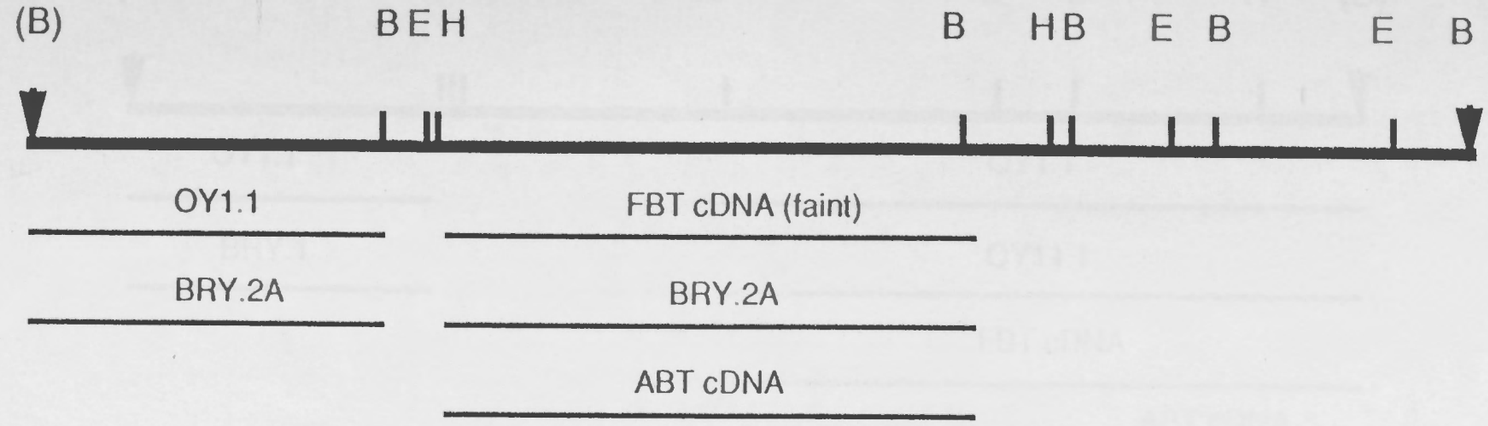


Figure 5.12. Restriction maps of EMBL3A.Y4 and EMBL3A.Y5, showing the fragments to which the various probes hybridize.

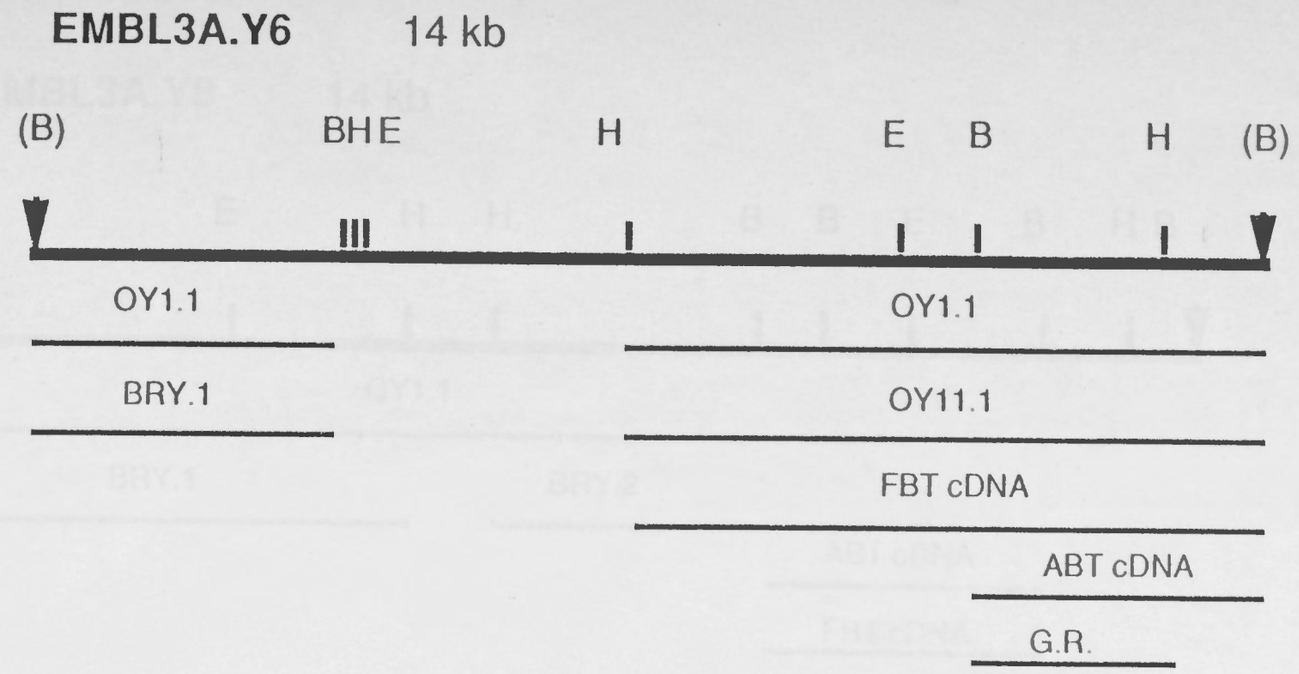


Figure 5.13. Restriction maps of EMBL3A.Y6 and EMBL3A.Y7, showing the fragments to which the various probes hybridize.

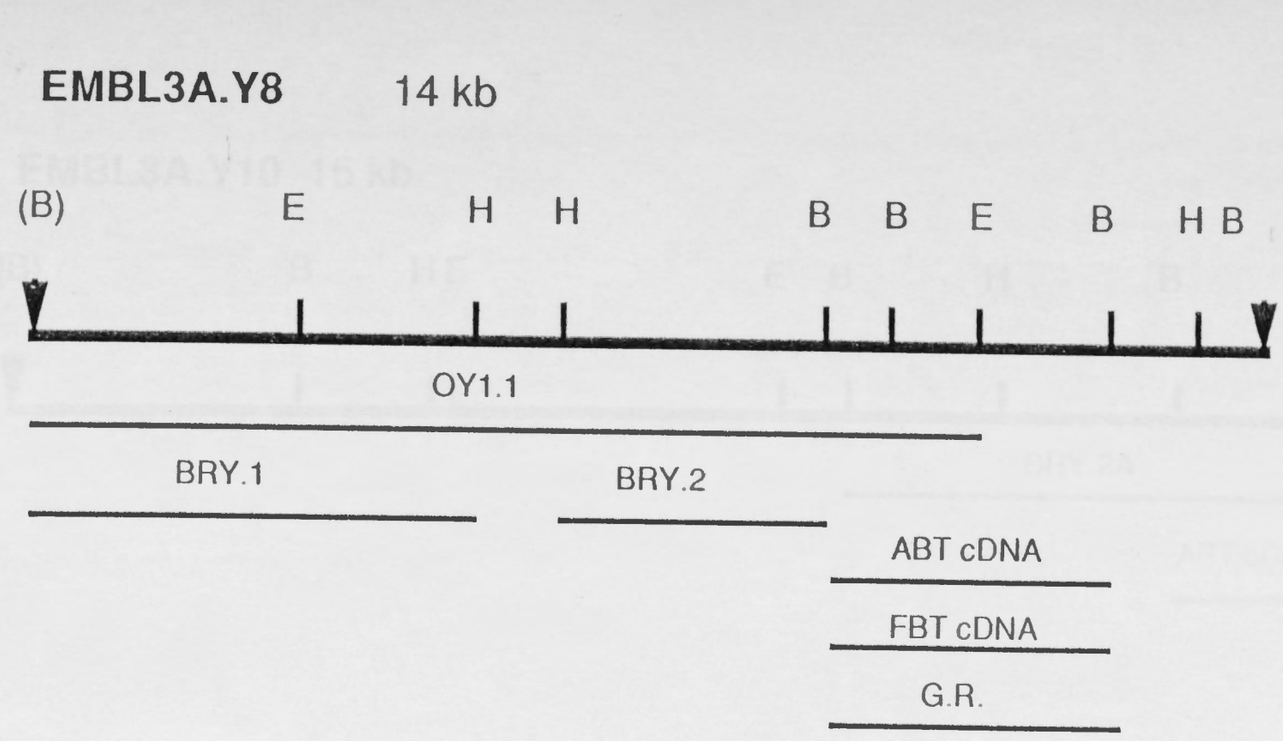


Figure 5.14. Restriction maps of EMBL3A.Y8 and EMBL3A Y9, showing the fragments to which the various probes hybridize.

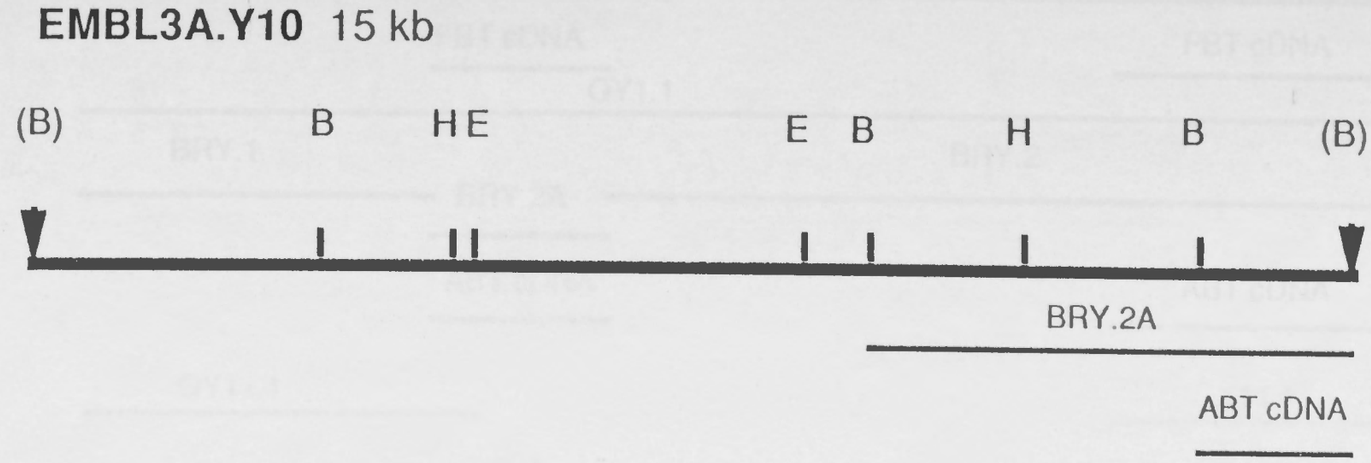


Figure 5.15. Restriction maps of EMBL3A.Y10 and EMBL3A.Y11, showing the fragments to which the various probes hybridize.

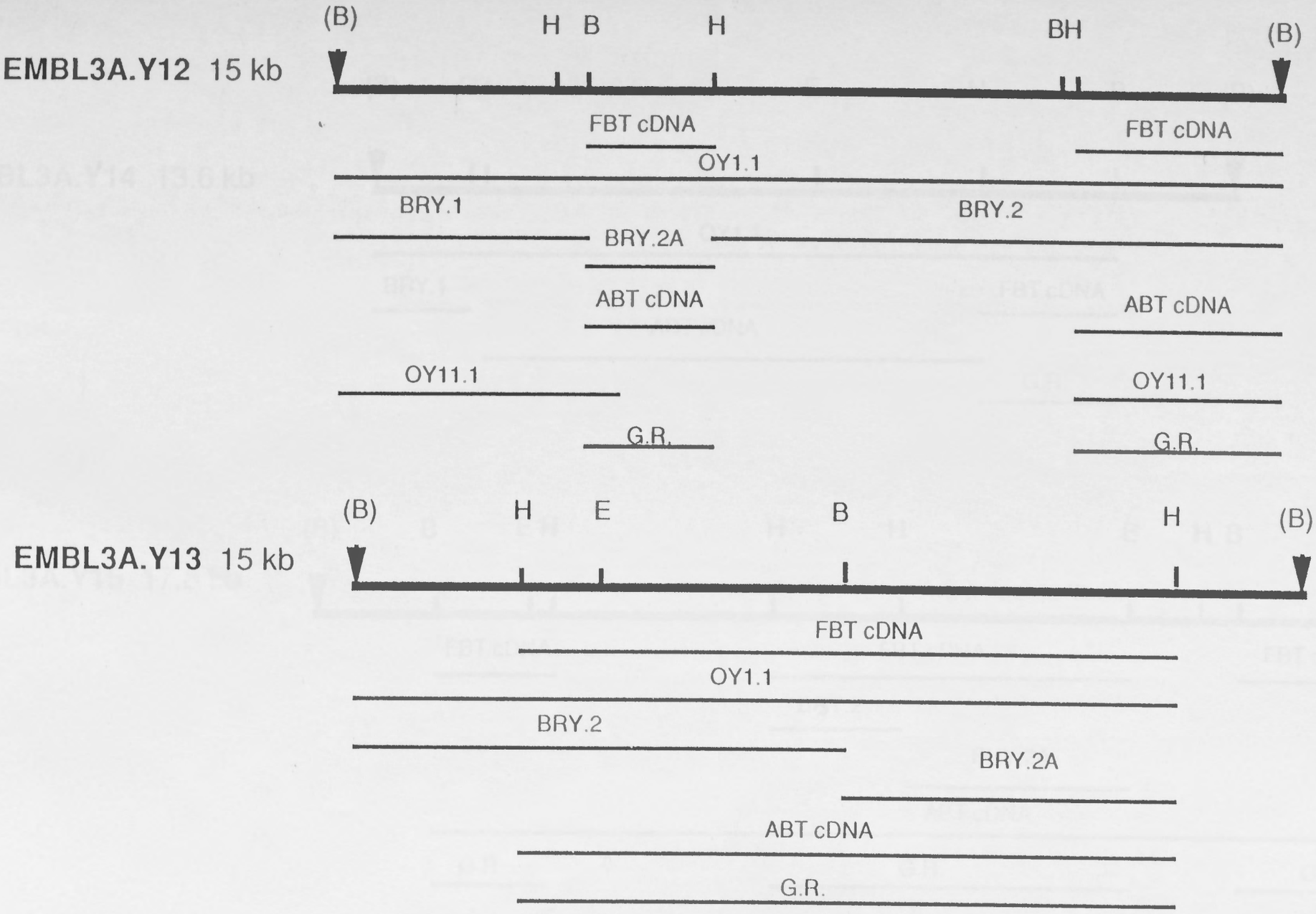
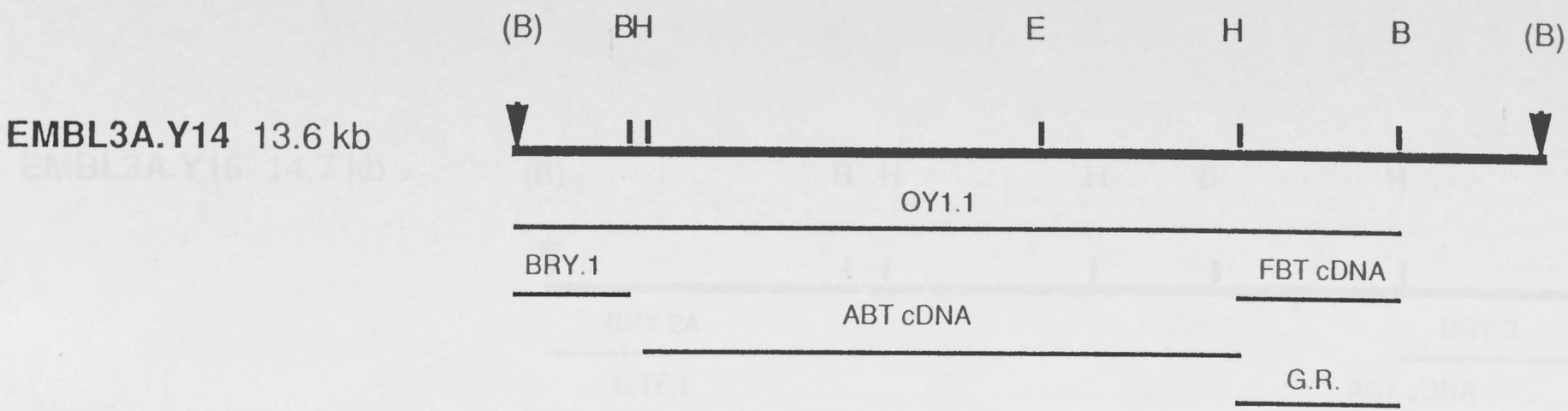


Figure 5.16. Restriction maps of EMBL3A.Y12 and EMBL3A.Y13, showing the fragments to which the various probes hybridize.



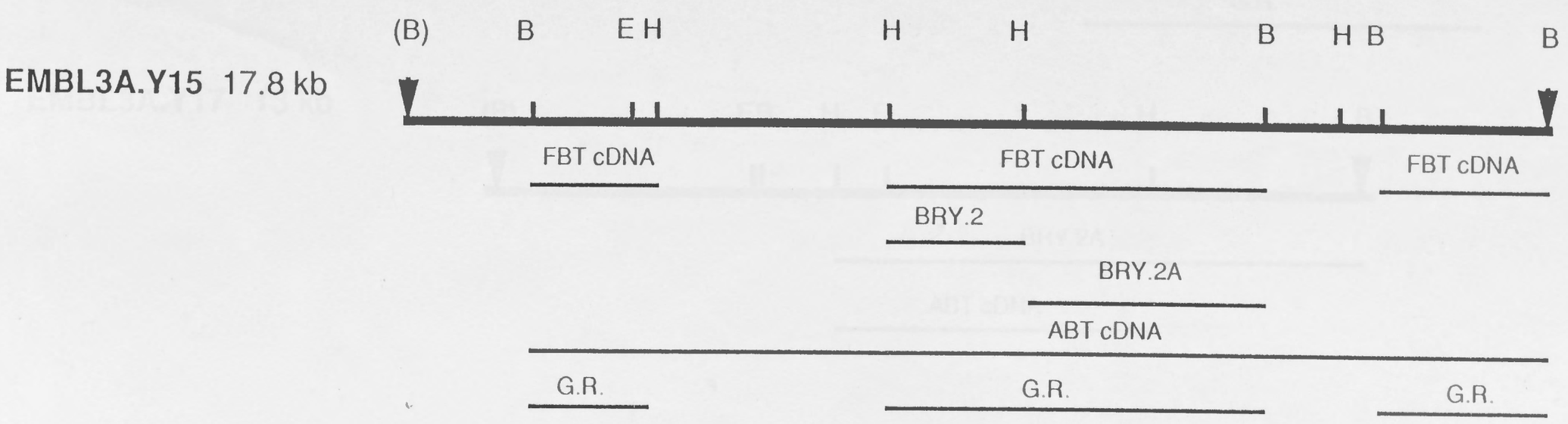


Figure 5.17. Restriction maps of EMBL3A.Y14 and EMBL3A.Y15, showing fragments to which the various probes hybridize.

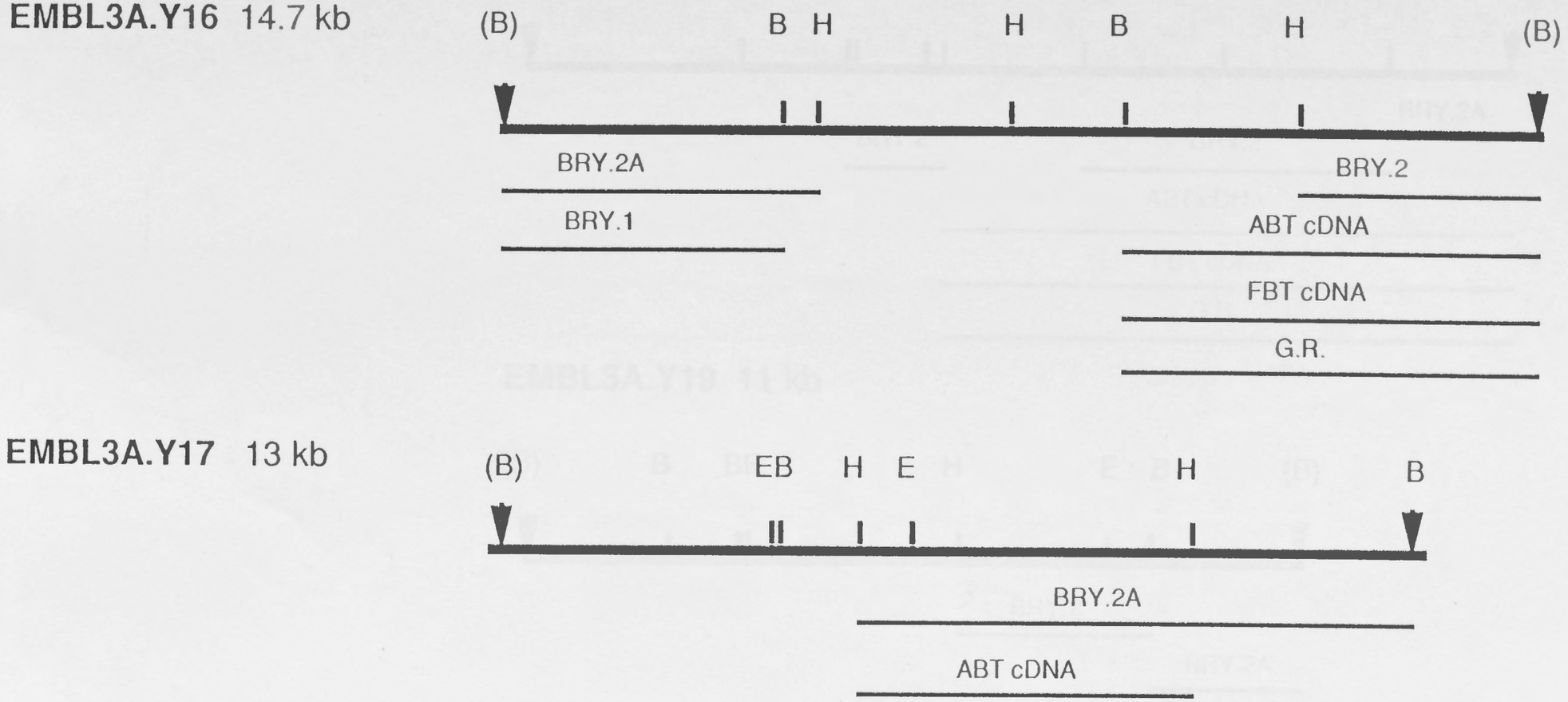


Figure 5.18. Restriction maps of EMBL3A.Y16 and EMBL3A.Y17, showing the fragments to which the various probes hybridize.

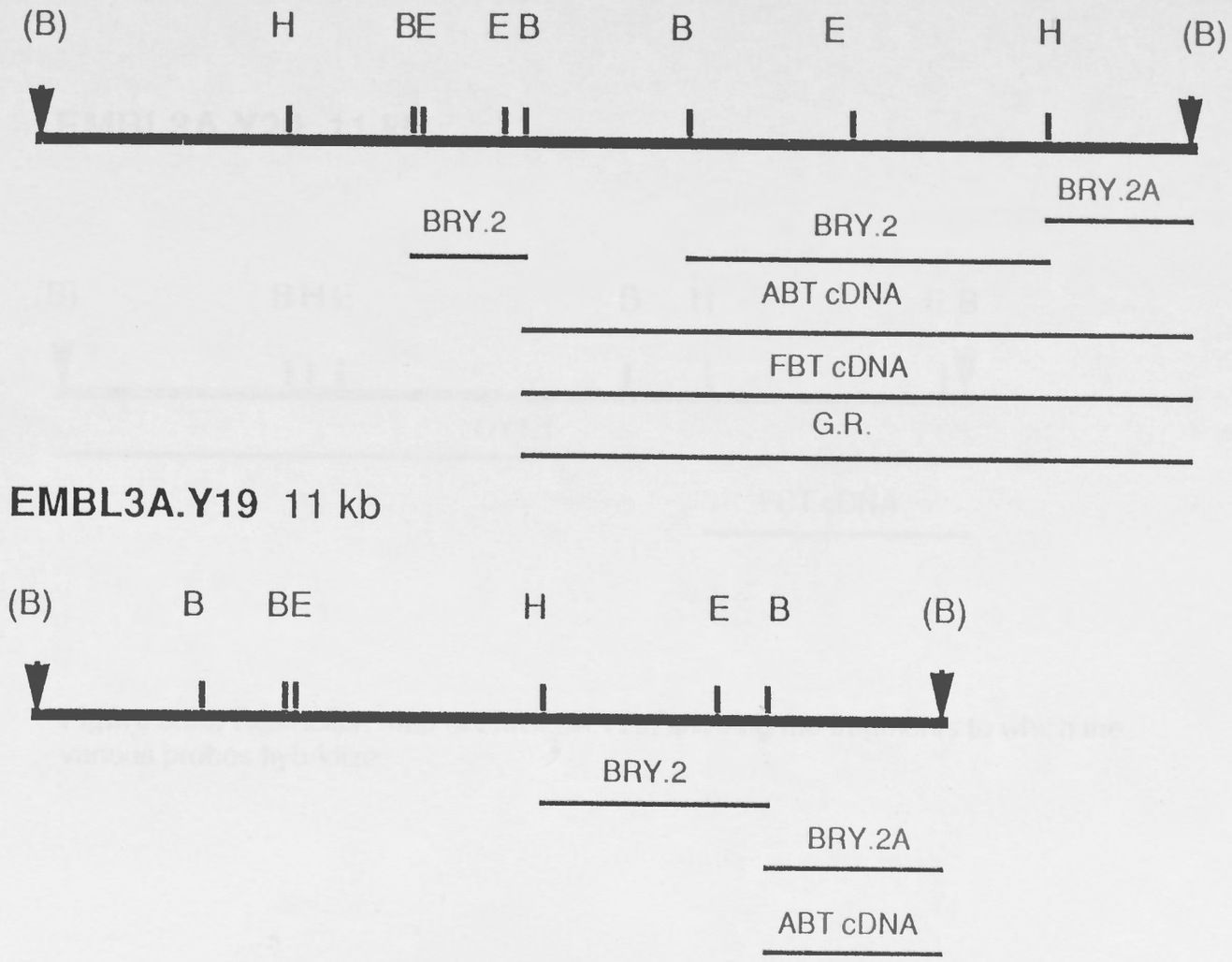


Figure 5.19. Restriction maps of EMBL3A.Y18 and EMBL3A.Y19, showing the fragments to which the various probes hybridize.

EMBL3A.Y20 11 kb

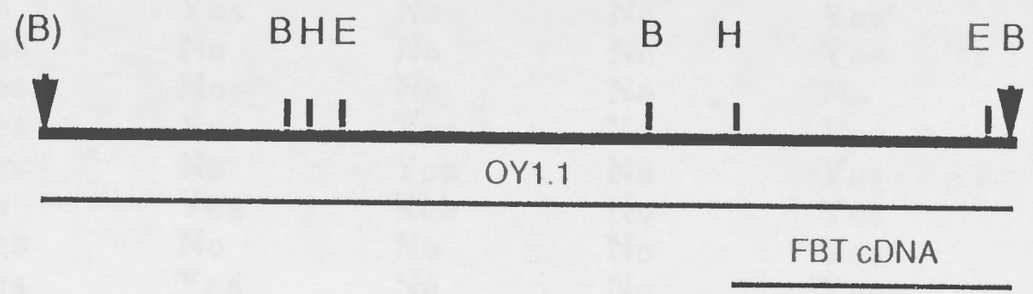


Figure 5.20. Restriction map of EMBL3A.Y20, showing the fragments to which the various probes hybridize.

Phage	BRY.2	BRY.2A	BRY.1	OY1.1	OY11.1	ABT cDNA	FBT cDNA	Female DNA
EMBL3A.Y1	Yes	Yes	Yes	Yes?	Yes	Yes	Yes	Yes
EMBL3A.Y2	Yes	Yes	No	Yes	Yes	Yes	Yes	Yes
EMBL3A.Y3	Yes	Yes	No	No	No	Yes	Yes	Yes
EMBL3A.Y4	Yes	Yes	No	Yes	No	Yes	Yes	No
EMBL3A.Y5	Yes	Yes	No	Yes	No	Yes	Yes	Yes
EMBL3A.Y6	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes
EMBL3A.Y7	Yes	Yes	No	Yes	No	No	No	No
EMBL3A.Y8	Yes	No	Yes	Yes	No	Yes	Yes	Yes
EMBL3A.Y9	Yes	No	Yes	No	No	Yes	Yes	Yes
EMBL3A.Y10	Yes	Yes	No	No	No	Yes	No	No
EMBL3A.Y11	Yes	Yes	No	No	No	No	No	No
EMBL3A.Y12	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
EMBL3A.Y13	Yes	Yes	No	Yes	No	Yes	Yes	Yes
EMBL3A.Y14	Yes	No	Yes	Yes	No	Yes	Yes	Yes
EMBL3A.Y15	Yes	Yes	No	No	No	Yes	Yes	Yes
EMBL3A.Y16	Yes	Yes	Yes	No	No	Yes	Yes	No
EMBL3A.Y17	Yes	Yes	No	No	No	Yes	No	No
EMBL3A.Y18	Yes	Yes	No	No	No	Yes	Yes	No
EMBL3A.Y19	Yes	Yes	No	No	No	Yes	No	No
EMBL3A.Y20	No	No	No	Yes	No	No	Yes	No

Table 5.1. A summary of the results of the hybridization analysis with various probes reported in the Results section of Chapter 5. The probes are BRY.2, subcloned from EMBL3A.Y1 (Chapter 3), BRY.2A, a male-specific restriction fragment subcloned from BRY.2, BRY.1 (Chapter 2), OY1.1, a male-specific clone from a genomic sheep library, OY11.1, another male-specific clone from a sheep library which may be a pseudogene, Adult Bull Testes cDNA, (ABT cDNA), Foetal Bull Testes cDNA, (FBT cDNA), and nick-translated cow DNA, used to detect genomic repeats in the phage.

pattern of hybridization of the various male-specific repeat probes, cDNA probes and female DNA does not indicate the presence of a discernibly recurring sequence.

5.4 Discussion

The isolation of twenty different phage containing DNA sequences from the bovine Y chromosome on the basis of their homology with BRY.2 and/or OY.1.1 has revealed a great deal about the structure of repeats on this chromosome. The DNA cloned in these phage represents approximately 2% of the chromosome and yet there is no evidence of any overlap between the phage as judged by a comparison of the restriction maps of the phage or a comparison of the patterns of hybridization of BRY.2A, BRY.1, OY.1.1, OY.11.1 or adult or foetal bull testis cDNA. This implies that the sequences hybridizing to each of the probes are dispersed along the Y chromosome, interspersed with each other, and with repeats from elsewhere in the genome, in no discernible pattern.

The overall impression gained from this study is that all of the phage isolated contain an assortment of various elements, those repeated only on the Y chromosome and those repeated elsewhere in the genome, many of which are homologous with sequences transcribed into poly(A+) RNA in the foetal and/or adult bull testis. The particular subset of male-specific repeated sequences found in each group of sequences (as defined by their presence in a single contiguous insert), though related to one another, varies considerably and there does not seem to be any indication of a constant order or arrangement of sequences, nor any indication of a tandem repeat structure. BRY.2 repeats may occur without BRY.1 sequences, as expected from the higher copy number of BRY.2 compared to BRY.1 (Chapter 3) and confirmed by hybridization to the various phage. A similar result was obtained from hybridization of BRY.1 and BRY.2 to sheep genomic clones (Lord, 1989), confirming the extremely heterogeneous nature of these Y-chromosomal repeats. The close association of sequences similar to the two sheep subclones OY.1.1 and OY.11.1 in the same recombinant phage was also found for recombinant phage from the sheep Y chromosome (Lord, 1989) and may be evidence of an ancient underlying regular repeat structure.

The fact that none of the phage isolated overlap may indicate that there are stretches of genomic repeat sequences between each group of Y-specific repeated sequences sufficiently long to prevent the cloning and selection for a phage with Y-specific repeated sequences at one end and dispersed genomic repeats at the other. This also implies that the stretches of predominantly male-specific sequences are short enough to be contained within an insert of 13 to 20 kb. There may also be a problem with not having a representative sample since only ~2% of the Y chromosome has been isolated while these sequences are estimated to comprise up to 40 % of the Y chromosome (Chapter 3).

It is difficult to envisage an original tandem structure from the very scrambled structure seen here, but it should be noted that an initial duplication event may have been followed by scrambling, perhaps due to the insertion of repeats from elsewhere in the genome, followed by a further duplication of these scrambled elements and so on, so that elements within the group were replicated in varying numbers and with varying surrounding sequences. The comparison of sheep and cattle sequences in Chapter 4 would seem to support this hypothesis.

Repeat units of a tandem array are often nearly identical (Card *et al.*, 1982), but polymorphism between individual repeat units in a tandem array is also common and has been documented (Boseley *et al.*, 1979; Korn and Bogenhagen, 1982), particularly for the human Y chromosome (Smith *et al.*, 1987). Bernstein *et al.* (1985) suggested the rodent and chicken U1 and U2 RNA genes might be part of an extremely polymorphic tandem array that has been severely scrambled by genetic exchange. Many tandemly repeated satellite sequences such as the primate alphoid families and even simple sequence satellites show polymorphism between repeat units (Singer, 1982b). Some polymorphisms leave the basic DNA sequence organization of the individual repeat units intact; these include variable numbers of an internally repetitious spacer sequence (Bosely *et al.*, 1979), or mutations such as single base changes and small insertions or deletions. Other polymorphisms alter the basic organization of the original repeat unit; these include larger insertions or deletions, for example the *Drosophila* histone gene repeat (Hentschel and Birnstiel, 1981), or homologous but unequal recombination between sequences that are present more than once within each individual repeat unit (Roberts *et al.*, 1983). In

fact polymorphisms consisting of variable numbers of a repetitious spacer sequence may arise from homologous but unequal recombination between fortuitous internal repeats within the basic repeat unit (Smith, 1976).

It is also possible that only parts of some elements on the bovine Y chromosome have been amplified due to their fragmentation from an original entire copy. The size of the groups of male-specific elements varies considerably between the inserts isolated, so that it is possible that in the future the sequences may become so scrambled and interspersed with genomic repeats that only very short, isolated Y-specific repeated elements will survive in a sea of genomic repeats on the cattle Y chromosome. Unequal sister chromatid exchange (Watson *et al.*, 1987) may have contributed to the amplification of these groups of sequences, perhaps initially as blocks of Y-specific repeated sequences and later, as repeats from elsewhere in the genome began to intersperse these elements, of blocks of Y-specific and non-Y-specific repeated sequences.

Many of the restriction fragments in the various phage to which the cDNA probes hybridize contain sequences which show homology with the nick-translated female genomic DNA probe. It is probably valid to assume that many of these fragments contain the highly repeated artiodactyl *Alu*-like genomic repeats (Watanabe *et al.*, 1982) which are known to achieve dispersal throughout the genome using reverse transcription and re-insertion into the chromosome (Sakamoto and Okada, 1985). Evidence for this includes the hybridization of cDNA to the region of EMBL3A.Y1 which hybridizes with the female repeat probe and which gives a highly repeated, dispersed pattern when used to probe a Southern blot of genomic DNAs (Chapter 3), and the presence of these repeats in bovine homologues to OY.11.1 (K. Matthaei, personal communication).

However the two *Bam* HI fragments BRY.2 and BRY.3 also hybridize with both foetal and adult testis cDNA probes and sequence analysis has shown that there are no artiodactyl repeats within these fragments, though they both hybridize with labelled female DNA, indicating the presence of other genomic repeats (Figure 5.10). Several of the other phage which show hybridization of specific restriction fragments with both the female probe and the cDNA probes (2, 3, 5, 6, 9, 12, 13, 14, 15) also give a hybridization

pattern showing the presence of a highly repeated, dispersed sequence when used to probe blots of genomic DNA (data not shown).

EMBL3A.Y8 is an exception, giving discrete bands of hybridization, including male-specific bands (Figure 5.2), though the map (Figure 5.14) shows the presence of a *Bam* HI restriction fragment that hybridizes to female DNA and both the cDNA probes. A male-associated pattern is probably still obtained because this phage insert is unusual in that it contains mostly sequences which are repeated specifically on the Y chromosome.

There are many different families of transcribed repeat sequences in the genomes of mammals, including sequences homologous to tRNA genes (Watson *et al.*, 1987) from which the artiodactyl SINEs may have evolved (Sakamoto and Okado, 1985), endogenous retroviruses (Tronick *et al.*, 1977) and the LINE families (Schmeckpeper *et al.*, 1984). The endogenous retroviruses are present as a multigene family with 10 to several 100 copies/ haploid genome (Benveniste and Todero, 1974). The type C virus isolated from deer is also present in the Bovidae, so that a transcribed repeat sequence found in all the artiodactyl species studied must be considered a candidate for membership of this family. Transcripts from members of the human LINE families are usually not polyadenylated (Schmeckpeper *et al.*, 1984), so if the same is true for LINEs from other mammals the sequences on the bovine Y chromosome to which the cDNA probes hybridize are probably not members of LINE families.

OY.11.1 may represent a pseudogene (Jacq *et al.*, 1977) which has a functional homologue on another chromosome (Lord, 1989) which, since its transfer to the Y chromosome (presumably by an RNA-mediated mechanism; Kole *et al.*, 1983), has been amplified on that chromosome. The hybridization of labelled cDNA to phage containing sequences homologous to OY.11.1 presumably reflects the continued transcription of the functional gene elsewhere in the genome.

It is however more difficult to understand the basis of hybridization of the cDNA probes to restriction elements which do not appear to contain highly repeated genomic sequences, but which do appear to contain Y-specific elements other than OY.11.1, as in phage numbers 4, 10, 17, 19

and 20. It is difficult to envisage how an RNA-mediated mechanism of replication could operate for sequences which are associated strongly with one chromosome. There is the possibility that these restriction fragments contain one copy of the *Alu*-like repeat which was not detected by the overnight exposure of the filter after probing with the nick-translated female probe, or that representatives of another, less highly repeated sequence are found within these fragments.

It is also possible that some of these Y-specific repeated elements, which do have homologous sequences in the female but at close to single copy, represent pseudogenes, like OY.11.1, which came to the progenitor ruminant Y chromosome and were then amplified, perhaps by sister chromatid exchange, and scrambled by interspersion with transcribed genomic repeats. The presence of open reading frames and short direct repeats in BRY.2 and BRY.3 may be evidence of such an origin, as may the homopurine tracts seen in BRY.3, which are often associated with genes (Christophe *et al.*, 1985).

It has been estimated that up to 20% of the mammalian genome may have arisen by reverse transcription of RNA and subsequent re-integration of the cDNA into the genome (Van Arsdell *et al.*, 1981; Bernstein *et al.*, 1983). Many genes which give rise to pseudogenes are expressed in germ line cells (Vanin, 1984). The human Y chromosome has pseudogenes for the housekeeping genes argininosuccinate synthetase and actin (Daiger *et al.*, 1982; Heilig *et al.*, 1984). The essentially non-functional nature of the Y chromosome may explain the easy re-integration of sequences into this chromosome (Leroy *et al.*, 1987).

Several phage show hybridization of the foetal bull testis cDNA probe to different restriction fragments than hybridize with the adult bull testis cDNA probe, for example 3, 5, and 14, but some of the phage hybridize only to one of the cDNA probes (only to foetal bull testis cDNA: 20; only to adult bull testis cDNA: 10, 17 and 19). It is also very intriguing to note that these phage appear to contain few sequences which are repeated elsewhere in the genome and the cDNA probe hybridizes with a restriction fragment containing Y-specific repeated sequences. Posakony *et al.* (1983) found that repetitive sequences in sea urchin are transcribed and that this transcription may change depending on the developmental stage of the cell.

Some transcripts of repetitive sequences have been implicated in the splicing of hnRNA to mRNA (Krayev *et al.*, 1980). In mammals, although repeats are transcribed very efficiently into hnRNA, most of these transcripts never reach the cytoplasm, being degraded in the nucleus (Kramerov *et al.*, 1982). The existence of stage-specific transcripts is consistent with the hypothesis that some of the repeats play regulatory roles (Riggs and Taylor, 1987), as suggested by Davidson and Britten (1979). At this stage it seems more reasonable to assume that the hybridization of only adult or foetal bull testis cDNA to a fragment containing male-specific sequences indicates the origin of that sequence as a pseudogene whose functional counterpart (presumably on another chromosome) is transcribed in the adult or foetal bull testis. The restriction fragments of the three phage which hybridize only to adult bull testis cDNA also contain BRY.2A, which may indicate that at least this sequence, which now has many copies on the bovine Y chromosome may have a functional counterpart on the Y chromosome. However not every phage, or every part of a phage containing BRY.2A, also hybridizes with the adult bull testis cDNA probe. It is also possible that some phage contain an element which may or may not be associated with BRY.2A which has homology with a transcript in the cDNA probe.

The isolation of a cDNA clone with homology to one of these Y-chromosomal sequences may help to answer some of the questions concerning the nature of the transcribed sequences and their role in the evolution of the Y chromosome. Chapter 6 describes the results from a screening of a cDNA library.

CHAPTER SIX

THE ISOLATION OF A CDNA CLONE HOMOLOGOUS TO A Y-CHROMOSOMAL SEQUENCE

CHAPTER SIX

The Isolation of a cDNA Clone Homologous to a Y-Chromosomal Sequence

6.1 Introduction

Many of the repeated sequences on the bovine Y chromosome, both those sequences specific to the Y chromosome and those which are also found elsewhere in the genome, are transcribed into poly(A+) RNA (Chapter 5). Many repeated sequences are known to be transcribed (Crampton *et al.*, 1981) and many of these sequences use RNA-mediated methods of dispersal in the genome. The presence on the Y chromosome of sequences with homology to poly(A+) RNA and genomic repeats may indicate that these sequences have reached the Y chromosome via such a means. It may be that a sole copy of a dispersed genomic sequence such as the artiodactyl *Alu*-like repeat is responsible for the hybridization to cDNA probes of the cloned segments of Y-chromosomal DNA.

Alternatively, if it is the male-specific sequences which are hybridizing to the probe, then sequence analysis of a cDNA clone may reveal whether the Y-chromosomal copy has been processed or otherwise altered. This may indicate whether it originally migrated to the Y chromosome as a pseudogene, the original functional gene being located on the X chromosome or on an autosome. These sequences may have a high copy number on the Y chromosome with only one or two copies elsewhere in the genome. The hybridization of several of the phage, including EMBL3A.Y1, to the sheep probe OY.11.1 (Chapter 5) would seem to indicate that at least one pseudogene has been integrated into the Y chromosome. The isolation and characterization of a transcript homologous to one of these Y-chromosomal sequences may provide more definitive information about the nature of the sequence elements which have homology with poly(A+) RNA.

A cDNA library containing sequences complementary to the poly(A+) RNA from foetal bull testis was screened with one of the subclones from EMBL3A.Y1, BRY.3. BRY.3 has a higher copy number in the males of cattle, sheep and goats than in the females of these species, but it does not

give such a definite male-associated pattern of hybridization in cattle and sheep in comparison with BRY.2. The presence of genomic repeats within this subclone obscures the fact that male goats have a higher copy number than females because it does not appear to be male-associated in hybridization to Southern blots of goat DNA (Chapter 3, Figure 3.15). It is possible that some of these genomic repeats may have spread to the Y chromosome through an RNA-mediated mechanism, since BRY.3 hybridizes to both adult bull testis cDNA and foetal bull testis cDNA probes (Chapter 5). BRY.3 also contains open reading frames, some interesting stretches of simple sequence and some suggestion of insertion of sequences, for example short direct repeats and inverted repeats.

6.2 Materials and Methods

6.21 cDNA library construction and screening

A foetal bull testis cDNA library was constructed by Sandra Beaton in the plasmid pTZ18U. Poly(A$^+$) RNA was prepared as described in Chapter 5. The RNA (0.1 mg/ml) was denatured with 0.02 mg/ml oligo(dT) at 80°C for 90 sec then immediately chilled in ice-water. The template/primer solution was incubated at 42°C for 1 h with 50mM Tris-HCl, pH 8.3, 20mM DTT, 7.5mM Mg acetate, 1mM each of dATP, dGTP and dTTP, 0.2mM dCTP, 0.1 mCi/ml [α-^{32}P]dCTP and 1200 units/ml AMV reverse transcriptase. The reaction was stopped by addition of EDTA to 0.02M and SDS to 0.4%, extracted with an equal volume of equilibrated phenol followed by IAC extraction and the product precipitated with 2M ammonium acetate and ethanol. After centrifugation to recover the cDNA/RNA hybrid it was redissolved in TE then reprecipitated. The extent of first strand synthesis was followed by PEI-cellulose chromatography. For second strand synthesis the cDNA/RNA hybrid was recovered by centrifugation and redissolved in 50mM Tris-HCl, pH 7.5, to give a concentration of about 0.04 mg/ml. The template/primer solution was incubated in 20mM Tris-HCl, pH 7.5, 5mM $MgCl_2$, 10mM ammonium sulphate, 0.2mM each of dATP, dCTP, dGTP and dTTP and 0.05 mg/ml BSA with 30 units/ml of RNase H (Boehringer) and 125 units/ml DNA polymerase I (Boehringer) for 1 h at 12°C, followed by 1 h at 16°C. Another 15 units/ml of RNase H was added and the incubation continued at 16°C for a further 30 min. The progress of this reaction was also followed with PEI-cellulose chromatography.

The cDNA was blunt-ended by S1 nuclease (Pharmacia) digestion, according to the manufacturer's instructions, at the end of second strand synthesis. The reaction was stopped as above and extracted with equilibrated phenol and then with IAC. The solution was concentrated to 100 µl by repeated butanol extraction and excess butanol was removed by ether extraction, then the cDNA was reprecipitated twice. The ends were filled by Klenow polymerase (Pharmacia) with dNTPs prior to ligation into the *Sma* I site of pTZ18U. The ligation mix was transformed into *E.coli* MC1061.1 cells (obtained from Charles Rice, Caltech) by electroporation with a Bio-Rad Gene-Pulser, according to the manufacturer's instructions. The total complexity of the library was 100,000 independent colonies. The amplified library was frozen in 20% glycerol at -70°C.

An aliquot of the library (20,000 clones) was screened with the gel-purified 4.2 kb *Bam* HI insert BRY.3 labelled by nick-translation. Positives were scraped from the plates and frozen at -70°C in 20% glycerol or diluted and re-plated for further screening. After the third round of screening clones were isolated and 'minipreps' of the DNA prepared.

6.22 Northern blots

RNA was extracted from various frozen cattle tissues as described in Chapter 5. Northern blots were prepared as described in Chapter 2 except that after electrophoresis the RNA was transferred to Zeta-Probe membrane in 10mM NaOH for 6 h (Vrati *et al.*, 1987), then the membrane was rinsed in 2 x PE, 0.1% SDS.

6.3 Results

6.31 A cDNA clone homologous to BRY.3

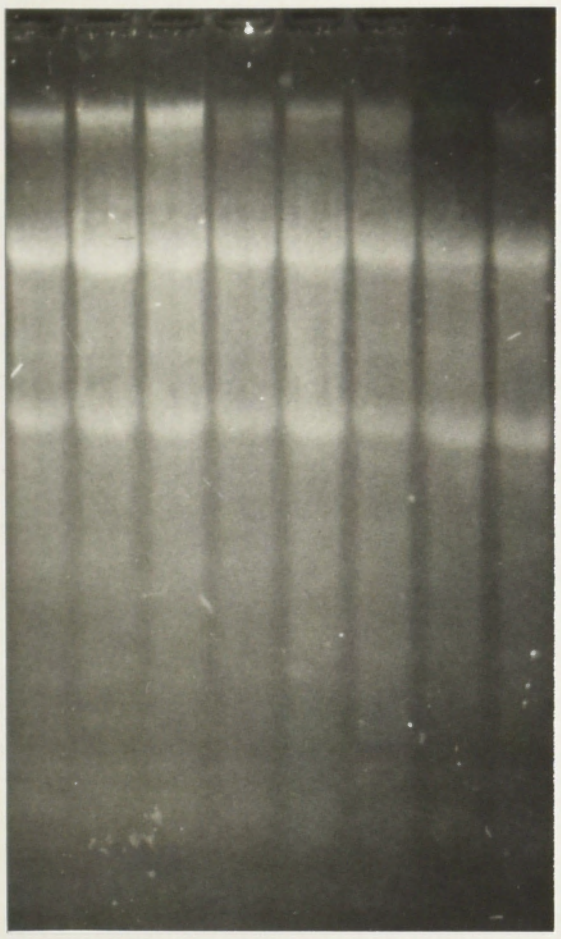
The primary screen yielded about 120 positive clones and ten of these were re-screened until three clones were isolated. Digestion of DNA from these three identical clones with several enzymes showed they contain an insert of approximately 200 bp of cDNA synthesized from foetal bull testis mRNA. The level of hybridization of sequences contained in BRY.3 to the cDNA library (approximately 0.6% of the clones screened) was unexpectedly high, and it was of interest that only one size of cDNA clone was recovered from the library, implying that only one sequence within BRY.3 is transcribed into poly(A+) RNA in foetal testis.

The plasmid clone, designated cBRY.3, was labelled by nick-translation and used to probe a Northern blot of RNAs from several different male and female cattle tissues. The gel was stained for visualization with acridine orange and the colour and staining of the RNA (Figure 6.1A) showed the ribosomal RNA species to be reasonably intact, indicating there was little degradation of the RNA. In Figure 6.1 it can be seen that the cDNA clone hybridizes to a single band of 2.2 kb RNA from adult bull testis, judging from the position of the 18S ribosomal RNA which is about 2.4 kb (Affara *et al.*, 1989). There was also hybridization to a high molecular weight RNA fraction from foetal and calf testis, and possibly to RNA from male spleen but there was no appreciable hybridization to RNA from female tissues. The level of transcription of this sequence is obviously higher in adult testis although this clone was isolated from a foetal calf testis cDNA library.

An identical Northern blot of RNAs was probed with BRY.3 labelled by nick-translation. Figure 6.2 shows that BRY.3 hybridizes much more strongly to RNA from adult testis than to any other tissue, with a major signal at about 2.2 kb, and other minor bands from the origin to about 2.6 kb. The probe hybridizes fairly strongly to high molecular weight RNA from the other male tissues: foetal and calf testis and spleen. There is not much hybridization to RNA from male liver or from ovary, female spleen or female liver. There is also some hybridization to the 18S and 28S bands in the RNA from all of the tissues examined. A long exposure (10 days) was needed to produce the signal with cBRY.3, compared to only 3 days for the hybridization with BRY.3, and even allowing for lower incorporation of label into the shorter cDNA clone probe it seems probable that BRY.3 contains another sequence in addition to the cBRY.3 homologue which hybridizes to RNA.

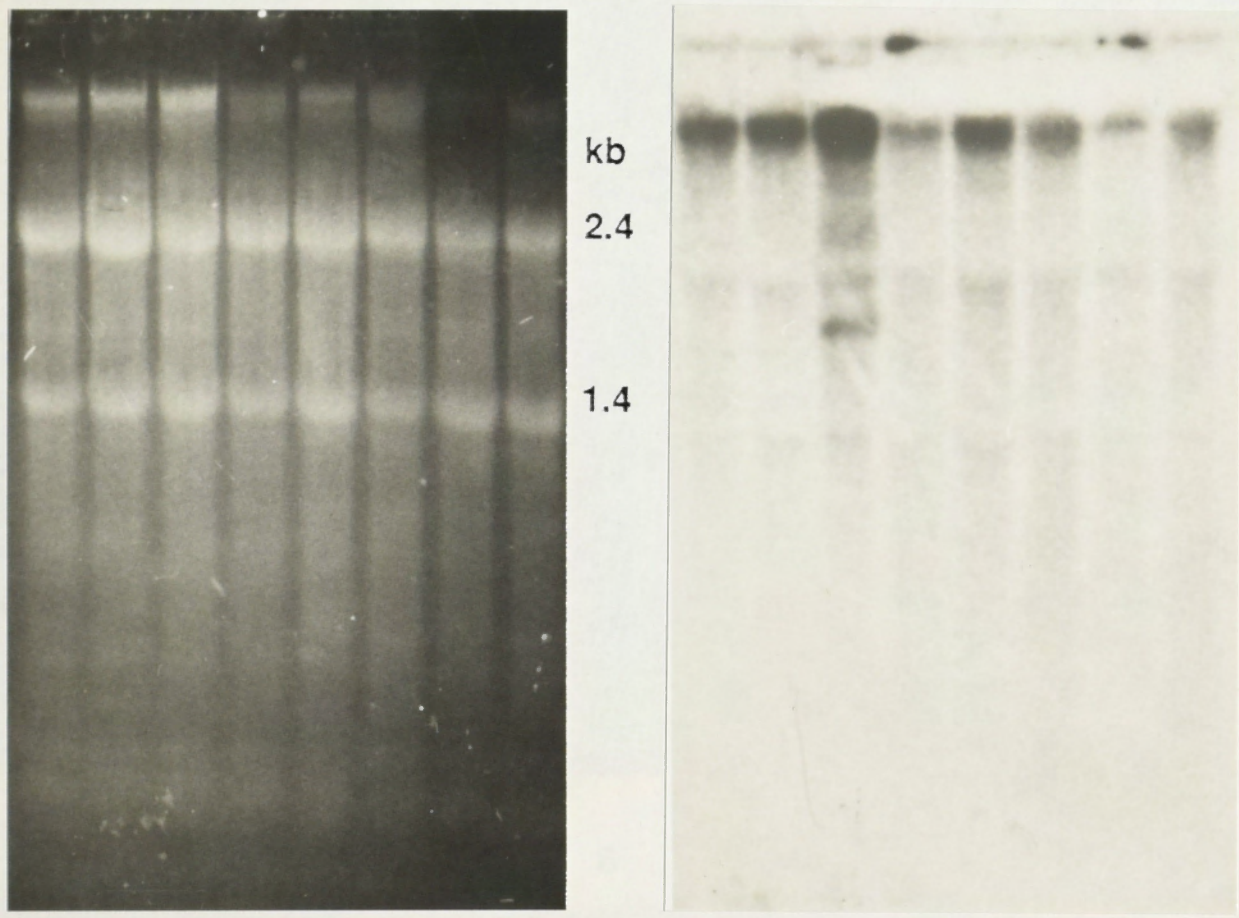
6.32 Southern blot analysis

A Southern blot of *Bam* HI digested DNA from the male and female of cattle, sheep, goats, deer and pigs was probed with nick-translated cBRY.3. A very interesting result was obtained (Figure 6.3). The cDNA clone appears to have a homologue in all the species examined, which is present at one or two copies in two *Bam* HI restriction fragments of about the same size (4.1 and 3.2 kb) in both sexes of each species. Bulls have the fragments seen in the cows, at the same intensity, but males also have



6.1A 6.1B

Figure 6.1. 10 µg of RNA from various cattle tissues: foetal testis (1), calf testis (2), adult testis (3), adult ovary (4), male spleen (5), female spleen (6), male liver (7) and female liver (8) was electrophoresed in 1.5% agarose as described in Materials and methods. The gel was stained with acridine orange for visualization (**6.1A**), and the sizes refer to the 18S (2.4 kb) and 28S (1.4 kb) ribosomal RNA species. The RNA was transferred to Zeta-Probe and the filter was hybridized with cBRY.3 which had been labelled by nick-translation. After 10 days exposure to XAR fim the autoradiograph (**6.1B**) was developed.

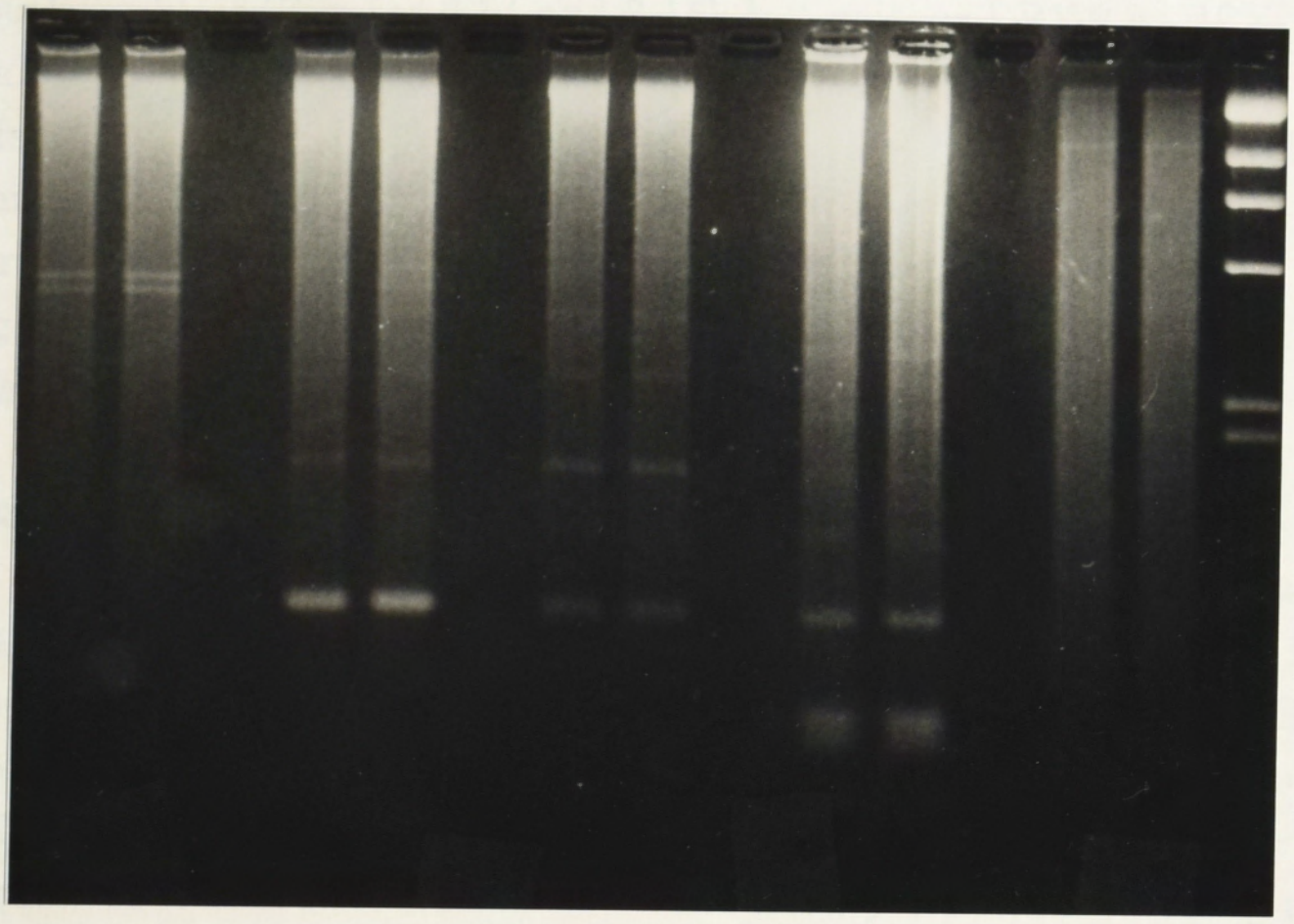


6.2A

6.2B

Figure 6.2. 10 µg of RNA from various cattle tissues: foetal testis (1), calf testis (2), adult testis (3), adult ovary (4), male spleen (5), female spleen (6), male liver (7) and female liver (8) was electrophoresed in 1.5% agarose as described in Materials and methods. The gel was stained with acridine orange for visualization (**6.2A**), and the sizes refer to the 18S (2.4 kb) and 28S (1.4 kb) ribosomal RNA species. The RNA was transferred to Zeta-Probe and the filter was hybridized with BRY.3 which had been labelled by nick-translation. After 10 days exposure to XAR fim the autoradiograph (**6.2B**) was developed.

6.3A



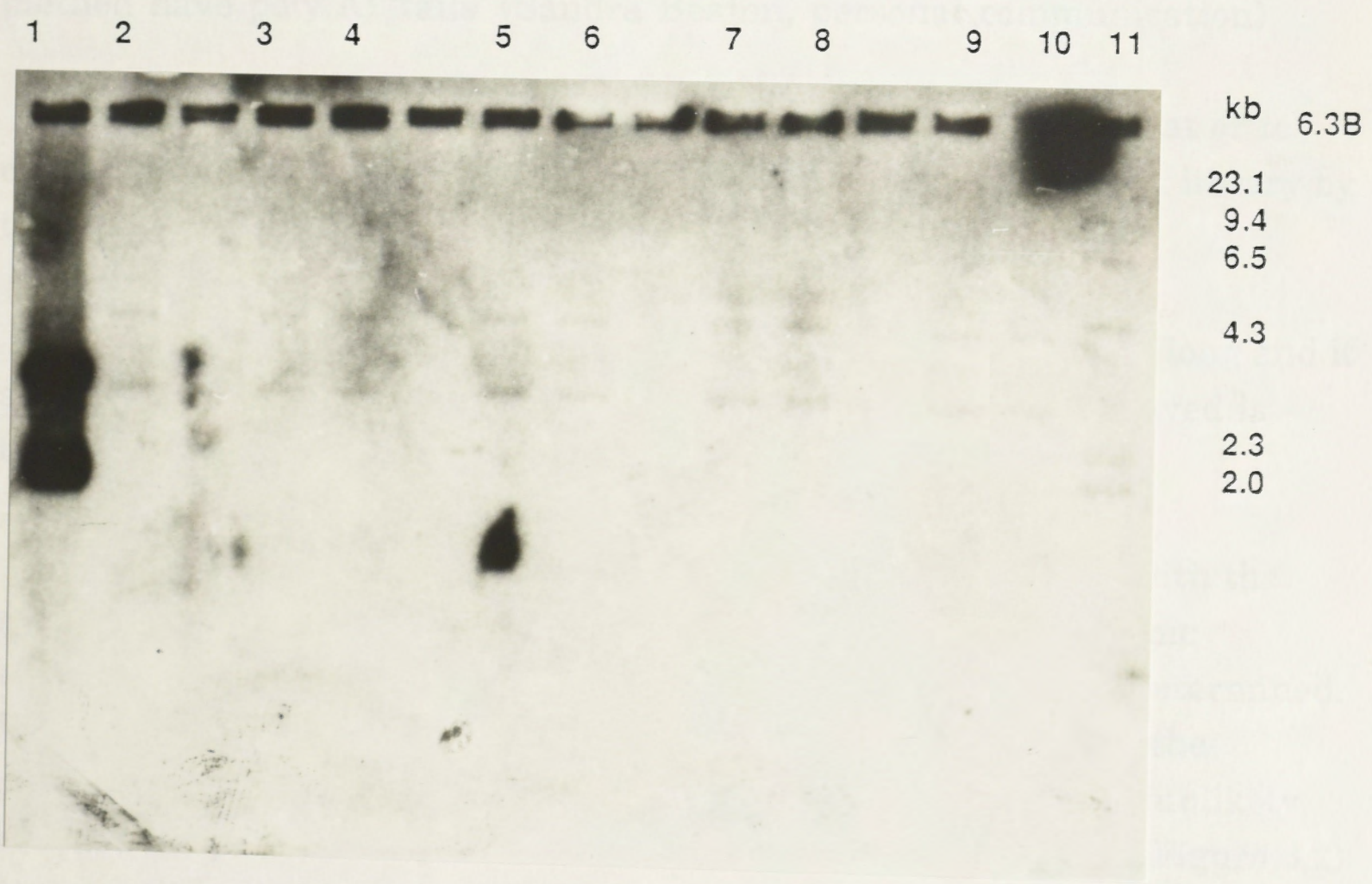


Figure 6.3. 2 µg of cattle male and female DNA (lanes 1 and 2), sheep male and female (lanes 3 and 4), goat male and female (lanes 5 and 6), deer male and female (lanes 7 and 8) and pig male and female (lanes 9 and 10) DNAs were digested with *Bam* HI and electrophoresed overnight at 12 volts in a 1% agarose gel, (**6.3A**). The DNAs were transferred by alkaline blotting to Zeta-Probe membrane and probed with nick-translated BRY.2, (**6.3B**). Lane 11 contains λ *Hind* III markers.

many more copies of this sequence in *Bam* HI fragments of 3.5 and 2.1 kb. All of these bands are also seen in hybridization of BRY.2 and BRY.3 to Southern blots (Figures 3.14 and 3.15), but neither BRY.2 or BRY.3 gives a male-specific pattern in only one species. It is clear however that though this sequence may be one of only a small number of sequences in BRY.3 which have homology with RNA, that cBRY.3 homologues are not the only non-Y-specific sequences which are present in BRY.3, since BRY.3 gives a high level of dispersed hybridization to *Bam* HI digested DNA in both sexes of all five species examined (Figure 3.15, Chapter 3).

6.33 Sequence of cBRY.3

The sequence of the cDNA clone was determined as described previously (Chapter 2) and is shown in Appendix 4. The sequence is 189 bp long and begins with a stretch of simple sequence DNA, $(AT)_7.(AG)_{17}$. There is no poly(A) 'tail', probably due to the method used to blunt-end the cDNA prior to cloning into the Sma I site of the plasmid. No other sequences isolated from another cDNA library constructed using the same method have poly(A) 'tails' (Sandra Beaton, personal communication).

The sequence shows no homology with the artiodactyl repeat or to cDNAs isolated by Sandra Beaton from the adult bull testis cDNA library by their homology with OY.11.1.

The longest open reading frame (start to stop) is only 18 nt long and it seems unlikely the mRNA from which this cDNA clone was derived is translated.

The cDNA clone was not found to have sequence identity with the sequence for BRY.3. It is possible that a copy of the cDNA genomic sequence is within the region for which sequence has not been determined. It is also possible that the clone was isolated using BRY.3 due to the presence of (GA) repeats in both sequences, however this is very unlikely because the pattern of hybridization of BRY.3 to a Northern blot (Figure 6.2) gives an almost identical pattern to that obtained when cBRY.3 is hybridized with a Northern blot (Figure 6.1). The cDNA clone also hybridizes to a 4.2 kb *Bam* HI fragment in both sexes of all the species examined which is the same size as one of the bands to which BRY.3 hybridizes. Finally it would be very coincidental if a fortuitously isolated

cDNA sequence showed such amplification on the cattle Y chromosome (Figure 6.3).

A comparison of the sequence of BRY.2 with cBRY.3 revealed no sequence identity, indicating that this sequence is not one of the poly(A+) RNAs with which BRY.2 hybridizes.

6.4 Discussion

The isolation of cBRY.3, a short cDNA sequence with no open reading frame which is homologous with BRY.3 but which has been amplified on the Y chromosome only in cattle raises more questions than it answers. The homologous sequence within BRY.3 is obviously not one of the non-Y-specific sequences within this subclone which lead to the dispersed pattern of hybridization of BRY.3 to goat and deer male and female DNAs, since it has been amplified only on the cattle Y chromosome. It was expected that a cDNA sequence which had homology with BRY.3 would represent transcription of a sequence which is spread throughout the genome by reverse transcription of a poly(A+) RNA.

It is possible that the cDNA clone isolated has not been transcribed from a Y-chromosomal copy of the sequence but from a functional copy elsewhere in the genome. The lack of an open reading frame in the cDNA may simply reflect the fact that a very truncated version of the poly(A+) RNA has been transcribed into cDNA in the clone recovered, judging from its hybridization to a large RNA species (Figure 6.1).

The part of the poly(A+) RNA which has been recovered in the cDNA clone may be the part which would later be processed out during the formation of mature mRNA, since the addition of poly(A) occurs before final processing (Crampton *et al.*, 1981). Long interspersed transcripts constitute a major fraction of the poly(A+) RNA stored in echinoderm and amphibian eggs (Richter *et al.*,1984) and intergenic repeated sequences are often included in polyadenylated pre-mRNAs. Calzone *et al.* (1988) isolated non-translatable poly(A+) RNAs from sea urchin eggs which are developmentally regulated but do not seem to be part of pre-mRNAs. It was suggested that these sequences, which contain three pairs of direct repeats, may have been involved in the frequent genome rearrangements seen in the

region from which they are transcribed. Perhaps cBRY.3 has similarities with these sequences since, although the short clone sequenced does not contain direct repeats, it has homology with an area of the genome which appears to have undergone a great deal of rearrangement. BRY.3 contains several direct repeats (Chapter 4) which are hallmarks of retroposition (Soares *et al.*, 1985).

This sequence seems to have undergone a comparatively recent amplification on the Y chromosome of only one species in the bovids, since the divergence of cattle from other bovids 15 to 20 million years ago (Novacek, 1982). The presence of this sequence at close to single copy in all the other species examined may indicate it is subject to natural selection in its original location. The tissue and developmental specificity of transcription of this sequence may indicate the autosomal/X homologue has a role in spermatogenesis.

Only cattle show a Y-specific increase in the copy number of the cBRY.3 homologue and considering its differential tissue and developmental expression it is tempting to speculate that a copy of this sequence reached the cattle Y chromosome via an RNA-mediated mechanism. After being inserted into Y-specific repeats it was perhaps amplified as part of the entire repeat unit. It is also interesting to speculate on whether the transcription of this sequence in the adult testis may have led to its transposition. Retroposition of sequences has been postulated to occur in the germ line (Vanin, 1984). In mice there is a high level of expression in the male germ tract (Del Villano and Lerner, 1976; Lerner, 1976) and retroviral sequences comprise up to 3% of the murine Y chromosome (Phillips *et al.*, 1982). While the level of transcription in foetal testis is high (a large number of clones hybridizing to the BRY.3 probe in the primary screen), the strength of the signal on the Northern blot shows that the level of transcription is much higher in adult bull testis. This high level of transcription in the adult testis suggests further that this sequence may have reached the Y chromosome through integration of an RNA. Lee *et al.* (1983) have predicted that most, if not all, multigene families whose expression occurs in germ-line cells of higher vertebrates are likely to contain RNA-mediated pseudogenes and that the numbers of pseudogene sequences in any given gene family may depend on the level of expression.

A BRY.2 homologue on the sheep Y chromosome, OY.11, contains a possible pseudogene, OY.11.1 (Lord, 1989), which, since its transfer presumably by reverse transcription of an mRNA and insertion, has been amplified on the Y chromosome of cattle, sheep and goats. This sequence also hybridizes predominantly to RNA from adult testis, and has homology with a cDNA sequence from adult bull testis. Leroy *et al.* (1987) have isolated two probes from a human Y chromosomal library which detect mRNAs specifically expressed in the testis. There are autosomal copies in humans and in mice the sequences are not located on the Y chromosome. These workers concluded that the Y chromosomal copies are probably pseudogenes and that the mRNAs they detect may be involved in spermatogenesis. They suggested that as the RNA is expressed during spermatogenesis re-integration by a mechanism of retroposition in the germ line would be possible. Other examples of pseudogenes on the human Y chromosome are those for the housekeeping genes argininosuccinate synthetase (Daiger *et al.*, 1982) and actin (Heilig *et al.*, 1984).

The presence of the homopurine tract $(AG)_{17}$ in the cDNA clone, may lend support to the notion that the sequence has originated from a coding sequence. Christophe *et al.* (1985) found a 209 bp long homopurine GA sequence between positions -512 to -304 of the human thyroglobulin gene. AGGAAAGAA is found 6 times in the sequence, suggesting the sequence originated from sequential duplication events. Similar homopurine sequences have been found in the vicinity of a number of genes, mostly in the 5' flanking region. In some cases they also have been shown to be repeated several times throughout the genome (Maroteaux *et al.*, 1983). The GA sequence may form a triple-helix DNA structure which could be involved in the control of gene expression (Christophe *et al.*, 1985).

The fact that this region of simple sequence is similar in BRY.3 and the cDNA clone may also be important if these tracts are a general feature of genes. Alternatively the presence of this sequence may simply reflect the fact that this sequence has been duplicated, especially since the homologue in BRY.3 contains a homopurine tract of 51 bp which includes the sequence GGAGA six times (Chapter 4). These tracts have been found at the ends of LINE sequences in the rat genome and it has been suggested these sequences might have a role in the transposition of these sequences (d'Ambrosio and Furano, 1987).

The transcript may be from an already transposed (Y-chromosomal) copy. It has been assumed that pseudogenes become inactive as soon as they are inserted into the genome because the RNA polymerase II promoter needed is not included in the transcript (Vanin, 1984), although the preproinsulin gene I in rats and mice is an exception, a functional retroposed gene which carries its own promoter (Soares *et al.*, 1985).

It is also possible that passive transcription due to the presence of a neighbouring SINE could result in continued transcription of a pseudogene. Although the artiodactyl SINEs (Watanabe *et al.*, 1982) were not found within BRY.2 or BRY.3, the end of BRY.3 is adjacent, in the original phage from which it was subcloned, to sequences which are repeated in very high copy number and dispersed throughout the genome (Figure 3.6, Chapter 3). These sequences have been identified in other bovine Y-chromosomal sequences (K. Matthaei, personal communication) and in cDNA clones from an adult bull testis library (S. Beaton, personal communication). There is a precedent for passive transcription by a neighbouring SINE. Manley and Colozzo (1982) found that an *Alu* I repeated sequence provided an RNA pol III promoter which resulted in the *in vitro* transcription of a substantial portion of an adjacent long, moderately repeated sequence, producing an RNA of 2.3 kb.

It would be interesting to hybridize the cDNA clone to Northern blots of RNA from tissues of other species, to determine whether this level of expression is typical or is peculiar to cattle, since the amplification of this sequence has occurred only in one species. This may provide a clue as to whether the amplified copies are being transcribed. The lack of an open reading frame in the cDNA clone may indicate that transcription is not from a functional gene. It is known that pseudogenes accumulate mutations including insertions and deletions and may themselves be duplicated (Little, 1982; Bernstein *et al.*, 1985). Most processed pseudogenes have multiple genetic lesions which prevent the translation of any transcript to give a functional polypeptide (Vanin, 1984).

This cDNA is not the only sequence on the Y chromosome which has homology with a transcribed sequence. Several of the phage isolated using BRY.2 (Chapter 5), including EMBL3A.Y1, hybridize with the sheep probe

OY.11.1 which has homology with transcripts in adult bull testis. It is unknown whether the Y-chromosomal sequences in the several other recombinant phage which hybridize with testis cDNA probes (Chapter 5) are related to cBRY.3. Hybridization of the cDNA clone to the restricted phage may provide further information. There is some indication from the hybridization of BRY.3 to a Northern blot (Figure 6.2) that BRY.3 has homology with another RNA species. This other RNA may not be polyadenylated or it may have not been recovered as a cDNA clone due to its low representation in comparison with cBRY.3 related clones.

These other transcribed sequences may be typical of the postulated non-Y-specific 'scrambler' sequences which have been inserted into blocks of Y-specific sequences and may be evidence of the proposed method of evolution of the structure of the cattle Y chromosome. As noted above though, this sequence is not present at a high enough copy number in female cattle or in either sex of any of the other species examined to account for the high level of hybridization of BRY.3 to these DNAs which is seen in Figure 3.15. BRY.3 obviously contains sequences which are present in quite high copy numbers elsewhere in the genome and which do not have homology with transcribed sequences as seen from the Northern blot of BRY.3.

The interspersion and scrambling of regularly repeated blocks of Y-specific sequences with transcribed genomic repeats has not been established by the isolation of this clone, however there is other evidence (Chapters 3, 4 and 5) that suggests this has also happened. The isolation of this clone implies that the Y-specific sequences may also originate through retroposition followed by amplification as part of a larger unit on the Y chromosome, as is the isolation of cattle cDNA clones with homology to a proposed pseudogene (Lord, 1989). The results of these experiments also further suggest that the cattle Y chromosome is undergoing more rapid change than is the sheep Y chromosome.

CHAPTER SEVEN

GENERAL DISCUSSION

CHAPTER SEVEN

General Discussion

7.1 The structure and organization of the bovine Y chromosome

When this work was begun almost nothing was known about the cattle Y chromosome at the molecular level. What information was available came from cytogenetic studies (Kurnit *et al.*, 1973) and related only to the lack of satellite DNA. The aim of this project was to isolate sequences repeated specifically on the Y chromosome of cattle which could be used to develop an assay for the sexing of pre-implantation embryos and which would allow access to the bovine Y chromosome for further study. The rationale for the project was the fact that Y-specific repeated sequences had been characterized for several other species (Cooke, 1976; Kunkel *et al.*, 1976; Lamar and Palmer, 1984), suggesting that the isolation of the Y chromosome from the rest of the genome because of its role in sex determination may make the accumulation of Y-chromosome specific repeats a general feature of this chromosome.

The construction of a deletion enrichment library (Lamar and Palmer, 1984) enabled the isolation of a short male-enriched sequence, BRY.1 (Chapter 2). All of the Y chromosomal repeats which have been isolated from other species have been repeated specifically on the Y chromosome in only one species or even one strain, however BRY.1 proved to be conserved as a Y chromosome-specific repeat in several species of the suborder Ruminantia.

At first it was thought this may be significant in terms of a possible association with coding sequences on the Y chromosome, however analysis of Y-chromosomal sequences surrounding BRY.1 in a recombinant phage clone (Chapter 3) indicated that the conservation is probably due to the extreme stability of the artiodactyl Y chromosome since the divergence of the various species. Indeed the general chromosomal stability of the Bovidae (cattle, sheep and goats) has been noted by cytogeneticists (Wurster and Benirschke, 1968). The isolation of a total of twenty non-overlapping phage genomic clones containing sequences from the cattle Y chromosome which include large blocks of sequences which are repeated elsewhere in

the genome repeats, as well as specifically on the Y chromosome (Chapter 5), confirmed that the Y-specific repeated sequences on the cattle Y chromosome are not present as one block but are distributed over at least 300 kb of the chromosome.

Once these sequences reached the Y chromosome in a progenitor ruminant there seems to have been an amplification event, leading to the presence of a much higher number of these sequences on the Y chromosome than elsewhere in the genome. The differences in copy number seen between the various species may simply indicate sequence divergence but the sequence identity of BRY.1 in cattle and sheep homologues (83%) may indicate the difference in copy number is due to continuing amplification and deletion events on the Y chromosomes of cattle, sheep, goats and deer. The quantitative polymorphism of BRY.1 in bulls (Chapter 2) may be further evidence of continuing change.

BRY.1 is present at close to single copy in both sexes of deer but it is found in BRY.2 which is male-enriched, at 75 copies in male deer and 35-40 in female deer. BRY.3, which also contains BRY.1, is present at 350 copies in both sexes of deer (Chapter 3). BRY.3 is not male-enriched in deer so the high copy number may simply reflect the presence of genomic repeats in this subclone, but BRY.2 does discriminate between male and female deer, indicating that BRY.1 has been amplified on the Bovidae Y chromosome after the original divergence of the Cervidae and Bovidae. Amplification of most male specific sequences seen in BRY.2 occurred before this divergence but after the divergence of the Suidae (pigs) from the ruminants.

The original male enriched sequences on the Y chromosome of the ruminant progenitor have also been broken up by the interspersion of sequences from elsewhere in the genome, as exemplified by the presence of varying numbers of genomic repeats in different phage inserts from the cattle Y chromosome. The isolation of clones from testis cDNA libraries which have homology with repeated sequences on the cattle (Chapter 6) and sheep Y chromosomes (Lord, 1989) suggests the means by which this interspersion has occurred.

The cDNA clone cBRY.3 probably also demonstrates the fact that the interspersion of Y-specific sequences with sequences from elsewhere in the

genome by RNA-mediated mechanisms is still occurring. This is further demonstrated by the fact that a bovine Y chromosomal homologue to OY.11.1 contains an artiodactyl repeat which is absent from the sheep clone (K. Matthaei, personal communication).

The continuing amplification of newly inserted sequences on the cattle Y chromosome is also demonstrated by the presence of a moderately repeated sequence on the cattle Y chromosome, which has homology to a poly (A+) transcript, while the DNA sequence is present in only one or two copies in cows and in both sexes of the other species examined. It is probable that sister chromatid exchange is contributing to the scrambling of the original structure of the Y chromosome, so that recently inserted sequences are being duplicated as a block with older Y chromosomal sequences, perhaps being broken up in the process themselves. BRY.1 is an example of a sequence element which is always associated with a larger block of other sequences when it occurs but which is not necessarily present in every block of similar sequences (Chapter 5). This may indicate that amplification of some representatives is leading to further divergence of sequence structure on the bovine Y chromosome. Amplification of repeated sequences on the cattle Y chromosome has resulted in these sequences comprising up to 40% of this chromosome (Chapter 3). The insertion of genomic sequences into the Y chromosomal repeats followed by replication of these altered blocks of sequences has resulted in extreme heterogeneity of sequence organization at the short range (10-15 kb) level.

There is some evidence from the sheep Y chromosome that the Y specific repeats shared by the ruminants did begin as tandemly repeated blocks, though no discernible pattern is obvious from characterization of cattle or sheep genomic clones. Lord (1989) used the CHEF (contour-clamped homogeneous electric field) system (Chu *et al.*, 1986) and rare cutting enzymes to obtain band patterns of 100-200 kbp intervals with hybridization of OY.1.1 to *Cla* I and *Sna* BI digests. The results may suggest a long range repeating unit defined by these enzymes on the ovine Y chromosome for OY.1.1.

In *Sfi* I digests a large fragment hybridizes to OY.1.1 at 450 kb with two clusters of stronger hybridization at approximately 200 and 350 kb. There is a close physical relationship between the OY.1 and OY.11 families

on the sheep Y chromosome, with both probes hybridizing to several of the genomic inserts of the phage isolated from the sheep library using BRY.2 (Lord, 1989). This is also seen for the cattle phage inserts (Chapter 5). Both OY.1 and OY.11 hybridized to similar discrete *Sfi* I fragments from digested sheep DNA, implying tandem organization of these sequence families and suggesting that these sequences together form a basic *Sfi* I repeating unit. The lowest molecular weight band was at about 40 kbp, increasing at a periodicity of 15-20 kbp up to 180 kbp. This would be evidence for the duplication of blocks of sequence after the insertion of a pseudogene sequence such as that in OY.11.1. OY.1.1 is present at a lower copy number than OY.11.1, perhaps reflecting its more recent acquisition by the Y chromosome, and therefore is not represented in every fragment containing OY.1.1, however this acquisition still occurred before divergence of the sub-families of the Bovidae as seen by the sequence homology between BRY.2 and OY.1 (Chapter 4) and by its presence in moderately repeated numbers in male cattle (Chapter 5).

Homology between OY.1 and BRY.2 extends further than the sequence element BRY.1 (Chapter 4), and sequence identity in the cattle and sheep homologues to OY.11.1 also extends past the pseudogene sequence (Lord, 1989), indicating this sequence was inserted into the Y chromosome sufficiently long ago to have been duplicated as part of a unit of Y specific sequences in both species. The cattle Y chromosome probably shares several features of the long range organization of these repeats with the sheep Y chromosome.

Sequence comparison of sheep and cattle sequences (Chapter 4) has shown that the short-range structure of the Y chromosomes in the two species does differ. Sequences which are contiguous in OY.1 are interspersed with other sequences in BRY.2 and BRY.3. While the various sheep homologues to BRY.2 are heterogeneous, there is some evidence that they may be more similar to one another than are the cattle BRY.2 homologues. Some of the male-specific elements comprising OY.1.1 are not found in BRY.2 and indeed are found in recombinant phage containing cattle Y-chromosomal sequences which have no homology with BRY.2, as shown by the isolation of EMBL3A.Y20 (Chapter 5). The isolation of this phage reinforced the results of sequence comparison between cattle and sheep Y-chromosomal sequences which suggested that sequences which

are contiguous on the sheep Y chromosome have been interspersed with other sequences on the cattle Y chromosome.

As discussed in Chapter 3 the cattle Y chromosome is larger than the sheep Y chromosome and the greater interspersion of repeats specific to the Y chromosome with sequences which are also repeated on other chromosomes, as well as the isolation of Y-specific sequences which are not male-enriched in sheep (Leonard *et al.*, 1987; Bondioli *et al.*, 1989) may suggest the cattle Y chromosome is undergoing more rapid change, or has in the past, than is the sheep Y chromosome.

At first there seems to be a contradiction: the ruminant Y-specific repeats have been confined to the Y chromosome due to both the special isolation of the Y chromosome from recombination with other chromosomes, which is a general feature of mammalian Y chromosomes and has lead to the presence of Y-specific repeats in most species, and the very stable chromosomal system of the artiodactyla which does not have many translocations or inversions of chromosome fragments, conversely the Y chromosomes in these species have been seen to be in an apparently continuous state of change. The conflict is resolved when it is understood that the Y chromosome is not isolated from those sequences in the genome which are transcribed and which can be copied into DNA and inserted into the Y-specific sequences. These inserted sequences may even promote recombination between copies of themselves on the Y chromosome, leading to further duplication and scrambling of the original structure.

7.2 Implications of this study for mammalian Y chromosome evolution

A better understanding of chromosome structure and function and evolution can be obtained by analyzing DNA sequences in different chromosomes. The results of this work support and extend the conclusions of several other authors concerning the evolution of the mammalian Y chromosome in particular and the evolution of genomes in general. Sequence acquisition and amplification are important in the evolutionary divergence of mammalian Y chromosomes (Eicher *et al.*, 1989). Studies of the Y chromosomes of humans and other primates (Cooke *et al.*, 1982; Kunkel and Smith, 1982) have shown that there has been integration of an autosomal sequence into the Y chromosome with subsequent amplification. The evolutionary duplication or amplification of portions of the Y

chromosome in several species has also been suggested by the finding that several Y chromosome-specific sequences occur in different regions of the Y chromosome (Vergnaud *et al.*, 1986).

The presence of the Y chromosome-associated sequences in similar numbers and with a similar long-range repeat structure on the Y chromosomes of cattle and sheep would seem to indicate the initial amplifications must have occurred fairly rapidly and lends support to the theory of saltatory amplification (McKay *et al.*, 1978). Secondary amplifications are known to occur at even higher frequency than that of the primary amplification event (Johnston *et al.*, 1983).

The internal heterogeneity of the Y-chromosomal repeats may be related to the model of an amplification gradient. Roberts *et al.* (1983) transfected cells in culture with a plasmid containing a selectable gene for ampicillin resistance. The sequence underwent gene amplification resulting in a tandem array of at least 20 units, joined together by homologous recombination between the repeats. The individual units within the array are heterogeneous in size and in sequence content, suggesting a model of sequence amplification involving cycles of DNA replication at a single locus, followed by multiple recombination events. The extent of sequence amplification is not equal throughout the unit- a gradient of amplification exists, with sequences at the centre of the unit being amplified more often than those at the periphery. The amplification of Y-chromosomal sequences seen on the bovine Y chromosome may reflect and support this model of amplification.

It has been suggested that the Y chromosome is vulnerable to colonization by repeated sequences, especially those which are RNA-derived (Craig *et al.*, 1987). The integration of RNA-derived sequences into the Y chromosome followed by their subsequent amplification has been seen for the retroviral related sequences on the mouse Y chromosome (Phillips *et al.*, 1982). Eicher *et al.* (1989) found that this insertion had occurred before the divergence of the subgenus *Mus* and that the subsequent amplification of the virus and its flanking cellular sequences occurred in the Y chromosome of the male progenitor to *M. musculus* and *M. domesticus*. In these two species the Y chromosome now contains about 500 copies of this sequence (Eicher *et al.*, 1989). Sequence analysis confirmed that after

integration the 3' and 5' long terminal repeats of the virus were able to evolve independently, and that after amplification each copy of the virus also evolves freely. An endogenous Y-chromosome specific sequence has also been cloned from the human Y chromosome (Silver *et al.*, 1987). There are only two copies of this sequence in the male genome and it seems the second copy of the viral *env* sequence has arisen by duplication of the first inserted sequence. The conservation of DNA 3' to both copies of the viral sequence would seem to support this view.

The retroposition of non-viral cellular RNA species has emerged as a major evolutionary force (Weiner *et al.*, 1986), leading to continuous sequence duplication, dispersion and rearrangement within eukaryotic genomes. The integration of pseudogenes into the Y chromosome has been well documented (Daiger *et al.*, 1982; Heilig *et al.*, 1984; Leroy *et al.*, 1987). Direct evidence for the continuing occurrence of this process on the bovine Y chromosome has been obtained by the isolation of a cDNA sequence whose Y-chromosomal homologue shows recent amplification on that chromosome (Chapter 6).

The retroposed SINEs can be important in the evolution of chromosomes, including the Y chromosome. Representatives of the artiodactyl SINE family (Watanabe *et al.*, 1982; Schimenti and Duncan, 1984) have been found in BRY.2 sheep homologues (Lord, 1989) and in bovine sequences recovered using OY.11.1 as a probe (K.Matthaei, personal communication). Ellis and Goodfellow (1989) found that there had been an insertion of an *Alu*,sequence into the pseudoautosomal region 225 bp from the boundary of the pseudoautosomal region and the Y-specific region of the chromosome. The *Alu* element disrupted pairing at the old boundary, preventing recombination in this 225 bp region. This sequence began to diverge, leading to the establishment of a new boundary at the beginning of the *Alu* sequence. The great apes have this insertion and use the same pseudoautosomal boundary but Old World monkeys lack it and may use the old boundary. The extent of pairing between the cattle X and Y chromosomes at meiosis has not been firmly established and nothing is known about the existence or otherwise of a pseudoautosomal region in cattle.

Simple sequence tracts are also found on the Y chromosomes of many species. The main component of the Bkm sequences is a GA^T_CA simple repeat (Epplen *et al.*, 1982). These sequences are not confined to the Y chromosome and do not discriminate between male and female in many species (Miklos *et al.*, 1989) but they are abundant on the mouse Y chromosome (Singh *et al.*, 1984). Organ-specific transcripts which are differentially regulated in various organs were found in mice. Although these RNAs contain open reading frames (Epplen *et al.*, 1983), no translation products of the sense-strand (TATC-TGTC) have been found (Epplen *et al.*, 1983). cDNA clones containing these sequences were found to also contain inverted repeats reminiscent of the inverted repeats of transposons (Epplen *et al.*, 1983). The transposon character of some members of the repeat family would explain the rapid spreading of the simple repeats in eukaryotic genomes. GA^T_CA repeats have also been implicated in recombinational 'hot spots' (Kobori *et al.*, 1986; Steinmetz *et al.*, 1987; Epplen *et al.*, 1988). It has been suggested (Epplen, 1989) that the GA^T_CA repeats may arise from a single point mutation in the abundant $(CA)_n$ repeats, which could then be spread through the genome by transposition. These tracts of $(CA)_n$ are also found on the cattle (Chapter 4) and sheep (Lord, 1989) Y chromosomes. The rapid amplification and contraction of these simple repeats would be especially facilitated in chromosomal regions with reduced crossover frequencies like the sex chromosomes (Epplen and McLaren, 1989).

One theory (Miyata *et al.*, 1987) proposes that there is a difference in the mutation frequency between the autosomes and each of the sex chromosomes because of the difference in the number of germ-cell divisions; sperm cells experience many more cell divisions than eggs (Albert *et al.*, 1983). The X-linked genes appear to be very conserved in comparison with autosome-linked genes (Miyata *et al.*, 1987) while the Y chromosome appears to be changing rapidly. They compared the human Y-linked argininosuccinate synthetase pseudogene and the chromosome 7 pseudogene. These pseudogenes are derived one from the other but not directly from the functional gene (Nomiyama *et al.*, 1986). The Y-linked pseudogene has changed at almost twice the rate of the autosomal pseudogene. Of course these differences may be due to the fact that sequences on the Y chromosome, except for the presumed few genes needed for sex determination and spermatogenesis, are not subject to the same

constraints and possible conversion pressures imposed on X and autosomal sequences.

Eicher *et al.* (1989) concluded that the Y chromosome may be evolving at a faster rate than other mammalian chromosomes because of the achiasmatic nature of most of the chromosome. Without the need to pair in a homologous manner with the X chromosome at meiosis this chromosome is free to gain, lose and amplify DNA sequences. The other Y chromosomes which have been studied show quite rapid change, as shown particularly by the lack of Y-specific sequences which are conserved on the Y chromosome of closely related species. This study of the bovine Y chromosome and comparisons with the Y chromosomes of other ruminant species has shown that this isolation from the rest of the genome may also lead to much greater conservation and confinement of repeated sequences on the Y chromosome, particularly in a stable chromosomal situation.

7.3 Embryo sexing

One of the reasons for studying the Y chromosome of cattle was the perceived economic advantage of being able to determine the sex of preimplantation livestock embryos. The isolation of BRY.2 (Chapter 3) allowed the isolation of OY.11.1 (Lord, 1989) which lead to the development of an assay for the presence of the Y chromosome in cattle, sheep, and goats (Matthews *et al.*, 1987).

Such an assay, to be economically practical, must be capable of being performed in the field and of giving same day results. Two 21-base oligonucleotide primers for a polymerase chain reaction (PCR) (Saiki *et al.*, 1985) assay have been derived from base sequences 120 bp apart, that are identical in cattle, sheep and goat OY.11.1 homologues (Matthaei and Reed, in preparation). These primers are used in an on-farm sexing assay which is very fast and accurate (Herr *et al.*, 1989a) and allows the same-day transfer of embryos of the desired sex into recipient cows. Day 7 embryos are flushed from the reproductive tracts of superovulated cows and a biopsy of each embryo is transferred to a tube containing the primers and reaction mix for the PCR assay. The reaction results in preferential amplification of sequences in the male embryos, resulting in a band on agarose gel electrophoresis in male samples which is absent from female samples (Herr *et al.*, 1989b). Of course the ideal would be to separate X and Y

bearing sperm fractions prior to fertilization, and the results of this work on the bovine Y chromosome may also facilitate that goal.

The ability to diagnose the presence of a Y chromosome has also found an application in compliance testing for the premium bull meat trade (Matthaei *et al.*, 1990). Meat samples are taken at random from stores and hybridized in a dot blot assay with Y-chromosomal probes and control satellite probes. The greater quality assurance possible with this test protects the value of meat products.

7.4 Further research

A great deal of work could be done to further characterize the sequences from the cattle Y chromosome which are contained in the 20 recombinant phage which have been isolated (Chapter 5). More sequence analysis is needed to establish whether some of these sequences are of retroviral origin, true processed pseudogenes or could have been carried to the Y chromosome as part of read-through SINE transcripts. Now that a clone containing one of the artiodactyl repeats first described by Watanabe *et al.* (1982) is available it would be useful to re-probe the phage inserts with this sequence in an attempt to determine how much of the hybridization with testis cDNA probes is due to the presence of these sequences. The same experiment could be repeated with the cDNA clone cBRY.3. In combination these experiments would give an indication of how many of these inserts have homology with other transcribed sequences. Isolation of a genomic clone, preferably from a female genomic library, containing the homologue of cBRY.3 might reveal whether this sequence is a functional message.

Further work is also required to establish whether the 40% of the cattle Y chromosome which is comprised of sequences related to BRY.2 and BRY.3 is concentrated in one region of the chromosome or distributed along its entirety. The small size of the cattle Y chromosome makes the use of conventional *in situ* hybridization to metaphase spreads of chromosomes for the mapping of repeated sequences to specific regions of the Y chromosome quite difficult. *In situ* hybridization to metaphase chromosomes from female cattle, sheep, goats and deer may help to determine which chromosome(s) BRY.1 originated from since it is present at only one or two copies in female cattle, sheep and goats and lacks sex

specificity in deer (Chapter 2). *In situ* hybridization may also be a useful tool for establishing whether the amplified transcribed sequence homologous to cBRY.3 has originated from the X or the autosomes. If sequences are reaching the Y-chromosome through RNA intermediates their functional counterparts will not necessarily be found on the X or Y chromosomes. However it is possible that an X-chromosomal copy of a sequence could be translocated to the Y during synapsis and then amplified, so that the new copies are still transcribed, or that a functional copy already on the Y chromosome could be amplified on that chromosome. A similar study for OY.11.1 would also be useful for comparison of results.

Extension of long range mapping techniques such as pulsed field gel electrophoresis (Schwartz *et al.*, 1982) to the cattle Y chromosome would be of interest for comparison with the results obtained for the sheep Y chromosome. The isolation of clones containing Y-chromosomal sequences from a library constructed in Yeast Artificial Chromosomes (YAC) (Burke *et al.*, 1987) will now be possible. As well as allowing precise study of the short to medium term organization of these repeats on the Y chromosome this approach has additional benefits in comparison to the other available avenues for mapping, in that the mapped DNA is recovered for further study.

It will be possible to select out a bank of clones from the cattle Y chromosome thus circumventing the need for flow sorting of Y chromosomes or cell hybrids containing the Y chromosome as the only cattle contribution as a source of enriched Y-chromosomal DNA. Duplicate screening of colonies with Y-chromosomal repeats and human and mouse pseudoautosomal probes may allow the identification of sequences from a cattle pseudoautosomal region.

The principle of deletion enrichment could also be applied to cDNA synthesized from the poly (A+) RNA from male and female tissues, to produce cDNA probes enriched for sequences expressed in the male. YAC clones from the bank containing Y-chromosomal sequences which hybridized with this probe could be selected for further study. As well as the possibility of retrieving further RNA-derived Y-chromosomal sequences, this approach may yield sequences which encode messages involved in the two vital functions of the Y chromosome; sex determination

and spermatogenesis. The most recent publications on the subject (Palmer *et al.*, 1989; Koopman *et al.*, 1989) present further evidence that *ZFY* is not the testis-determining gene *TDY*, therefore the challenge of isolating the sequence(s) on the Y chromosome which leads to male development remains.

REFERENCES

REFERENCES

Aaronson, S.A., Tronick, S.R. and Stephenson, J.R. (1976) Endogenous Type C RNA virus of *Odocoileus hemionus*, a mammalian species of new world origin. *Cell* **9**, 489-494.

Adams, J.W., Kaufman, R.E., Kretschmer, P.J., Harrison, M. and Nienhuis, A.W. (1980) A family of long reiterated DNA sequences, one copy of which is next to the human beta globin gene. *Nucleic Acids Research* **8**, 6113-6128.

Affara, N.A., Florentin, L., Morrison, N., Kwok, K., Mitchell, M., Cook, A., Jamieson, D., Glasgow, L., Meredith, L., Boyd, E. and Ferguson-Smith, M.A. (1986) Regional assignment of Y-linked DNA probes by deletion mapping and their homology with X-chromosome and autosomal sequences. *Nucleic Acids Research* **14**, 5353-5373.

Affara, N.A., Chambers, D., O'Brien, J., Habeebu, S.S.M., Kalaitsidaki, M., Bishop, C.E. and Ferguson-Smith, M.A. (1989) Evidence for distinguishable transcripts of the putative testis determining gene (*ZFY*) and mapping of homologous cDNA sequences to chromosomes X, Y and 9. *Nucleic Acids Research* **17**, 2987-2999.

Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K. and Watson, J.D. (1983) *Molecular biology of the Cell*. (Garland Publishing, New York).

Alvesalo, L. and de la Chapelle, A. (1981) Tooth sizes in two males with deletions of the long arm of the Y chromosome. *Annales HumanGenetics* **54**, 49-54.

Anderson, G.B. (1985) Identification of sex in mammalian embryos. In *Genetic Engineering of Animals, An Agricultural perspective.*, Eds. Evans, J.W. and Hollaender, A. (Plenum Press, New York) pp. 243-250.

Arnemann, J., Jakubiczka, S., Schmidtke, J., Schafer, R. and Epplen, J.T. (1986) Clustered GATA repeats (Bkm sequences) on the human Y chromosome. *Human Genetics* **73**, 301-303.

Ashley, T. (1984) A re-examination of the case for homology between the X and Y chromosomes of mouse and man. *Human Genetics* **67**, 372-377.

Barnes, W.M. (1977) Plasmid detection and sizing in single colony lysates. *Science* **196**, 180-182.

Benton, W.D. and Davis, R.W. (1977) Screening λgt recombinant clones by hybridization to single plaques *in situ*. *Science* **196**, 180-182.

Benveniste, R.E., and Todero, G.J. (1974) Multiple divergent copies of endogenous type C virogenes in mammalian cells. *Nature* **252**, 170-173.

Bernstein, L.B., Mount, S.M. and Weiner, A.M. (1983) Pseudogenes for human small nuclear RNA U3 appear to arise by self primed reverse transcripts of the RNA into new chromosomal sites. *Cell* **32**, 461-472.

Bernstein, L.B, Manser, T. and Weiner, A.M. (1985) Human U1 small nuclear RNA genes: extensive conservation of flanking sequences suggests cycles of gene amplification and transposition. *Molecular and Cellular Biology* **5**, 2159-2171.

Betteridge, K.J. (1989) Livestock sexing: past, present and future. In *Evolutionary mechanisms in sex determination,* Ed. Wachtel, S.S. (CRC Press, Boca Raton, Florida) pp. 279-289.

Birnboim, H.C. and Doly, J. (1979) A rapid alkaline extraction procedure for screening recombinant plasmid DNA. *Nucleic Acids Research* **7**, 1513-1523.

Bishop, C.E., Boursot, P., Baron, B., Bonhomme, F. and Hatat, D. (1985) Most classical *Mus musculus domesticus* laboratory mouse strains carry a *Mus musculus musculus* Y chromosome. *Nature* **315**, 70-72.

Bishop, C.E. and Hatat, D. (1987) Molecular cloning and sequence analysis of a mouse Y chromosome RNA transcript expressed in the testis. *Nucleic Acids Research* **15**, 2959-2968.

Bishop, C., Guellaen, G., Geldwerth, D., Fellous, M. and Weissenbach, J. (1984) Extensive sequence homologies between Y and other human chromosomes. *Journal of Molecular Biology* **173**, 403-417.

Blaho, J.A. and Wells, R.D. (1987) Left-handed Z-DNA binding by the recA protein of *Escherichia coli*. *Journal of.Molecular.Chemistry* **262**, 6082-6088.

Bondioli, K.R., Ellis, S.B., Pryor, J.H., Williams, M.W and Harpold, M.M. (1989) The use of male-specific chromosomal DNA fragments to determine the sex of bovine preimplantation embryos. *Theriogenology* **31**, 95-104.

Boseley, P.T., Moss, T., Machler, M., Portmann, R. and Birnstiel, M. (1979) Sequence organization of the spacer DNA in a ribosomal gene unit of *Xenopus laevis*. *Cell* **17**, 19-31.

Bostock, C.J., Gosden, J.R. and Mitchell, A.R. (1978) Localisation of a male-specific DNA fragment to a sub-region of the human Y chromosome. *Nature* **272**, 324-328.

Bouchard, R.A. (1982) Moderately repetitive DNA in evolution. *International Review of Cytology* **76**, 113-193.

Britten, R.J. and Davidson, E.H. (1969) Gene regulation for higher cells; a theory. *Science* **165**, 349-357.

Britten, R.J. and Davidson, E.H. (1971) Repetitive and non-repetitive DNA sequences and a speculation on the origins of evolutionary novelty. *The Quarterly Review of Biology* **46**, 111-138.

Britten, R.J., Graham, D.E. and Neufeld, B.R. (1974) Analysis of repeating DNA sequences by reassociation. *Methods in Enzymology* **29**, 363-405.

Britten, R.J. and Kohne, D.E. (1968) Repeated sequences in DNA. *Science* **161**, 529-540.

Brown, R.S., Sander, C. and Argos, P. (1985) The primary structure of transcription factor TFIIIA has 12 consecutive repeats. *FEBS Lett.* **186**, 271-274.

Brown, S.D.M. and Dover, G.A. (1981) Organization and evolutionary progress of a dispersed repetitive family of sequences in widely separated rodent genomes. *Journal of Molecular Biology* **150**, 441-166.

Brown, W.A.R. and Bird, A.P. (1986) Long-range restriction mapping of mammalian genomic DNA. *Nature* **322**, 477-481.

Brutlag, D.L. (1980) Molecular arrangement and evolution of heterochromatic DNA. *Annual Review of Genetics* **14**, 121-144.

Buckle, V., Mondello, C., Darling, S., Craig, I.W. and Goodfellow, P.N. (1985) Homologous expressed genes in the human sex chromosome pairing region. *Nature* **317**, 739-741.

Buhler, E.M. (1980) A synopsis of the human Y chromosome. *Human Genetics* **55**, 145-175.

Bull, J.J. (1980) Sex determination in reptiles. *Quarterly Review of Biology* **55**, 3-21.

Burgoyne, P.S. (1982) Genetic homology and crossing over in the X and Y chromosomes of mammals. *Human Genetics* **61**, 85-90.

Burgoyne, P.S. (1987) The role of the mammalian Y chromosome in spermatogenesis. *Development* **101** (Suppl.), 133-141.

Burgoyne, P.S., Levy, E.R. and McLaren, A. (1986) Spermatogenic failure in mice lacking H-Y antigen. *Nature* **320**, 170-172.

Burk, R.D., Szabo, P., O'Brien, S., Nash, W.G., Yu, L. and Smith, K.D. (1985a) Organization and chromosomal specificity of autosomal homologs of human Y chromosome repeated DNA. *Chromosoma* **92**, 225-233.

Burke, D.T., Carle, G.F. and Olson, M.V. (1987) Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* **236**, 806-812.

Burkhoff, A.M. and Tullius, T.D. (1988) Structural details of an adenine tract that does not cause DNA to bend. *Nature* **331**, 455-457.

Burns, J., Chan, V.T.W., Jonasson, J.A., Fleming, K.A., Taylor, S. and McGee, J.O'D. (1985) Sensitive system for visualizing biotinylated DNA probes hybridized *in situ*, rapid sex determination of intact cells. *Journal of Clinical Pathology* **38**, 1085-1092.

Calzone, F.J., Lee, J.J., Le, N., Britten, R.J. and Davidson, E.H. (1988) A long, nontranslatable poly(A) RNA stored in the egg of the sea urchin *Stronglocentrotus purpuratus*. *Genes and Development* **2**, 305-318.

Card, C.U., Morris, G.F., Brown, D.T. and Marzluff, W.F. (1982) Sea urchin small nuclear RNA genes are organized in distinct tandemly repeating units. *Nucleic Acids Research* **10**, 7677-7688.

Cathala, G., Savouret, J., Mendez, B., West, B.L., Karin, M., Martial, J.A. and Baxter, J.D. (1983) A method for isolation of intact translationally active ribonucleic acid. *DNA* **2**, 329-335.

Cattanach, B.M., Pollard, C.E. and Hawkes, S.G. (1971) Sex-reversed mice; XX and XO males. *Cytogenetics*. **10**, 318-337.

Cavalier-Smith, T. (1985) *The evolution of genome size*. (John Wiley and Sons, New York).

Chandra, H.S. (1985) Is human X chromosome inactivation a sex-determining device? *Proceedings of the National Academy of Sciences, USA*. **82**, 6947-6949.

Charnov, E.L. and Bull, J.J. (1977) When is sex environmentally determined? *Nature* **266**, 828-830.

Christophe, D., Cabrer, B., Bacolla, A., Targovnik, H., Pohl, V. and Vassart, G. (1985) An unusually long poly(purine)-poly(pyrimidine) sequence is located upstream from the human thyroglobulin gene. *Nucleic Acids Research* **13**, 5127-5144.

Chu, G., Vollrath, D. and Davis, R.W. (1986) Separation of large DNA molecules by Contour-clamped Homogeneous Electric Fields. *Science* **234**, 1582-1585.

Church, G.M. and Gilbert, W. (1984) Genomic sequencing. *Proceedings of the National Academy of Sciences, USA* **81**, 1991-1995.

Cooke, H.J. (1976) Repeated sequence specific to human males. *Nature* **262**, 182-186.

Cooke, H.J., Brown, W.R.A., and Rappold, G.A. (1984) Closely related sequences on human X and Y chromosomes outside the pairing region. *Nature* **311**, 259-261.

Cooke, H.J., Brown, W.R.A. and Rappold, G.A. (1985) Hypervariable telomeric sequences from the human sex chromosomes are pseudoautosomal. *Nature* **317**, 687-692.

Cooke, H.J. and Hindley, J. (1979) Cloning of human satellite III DNA: different components are on different chromosomes. *Nucleic Acids Research* **6**, 3177-3197.

Cooke, H.J. and McKay, R.D.G. (1978) Evolution of a human Y chromosome specific repeated sequence. *Cell* **13**, 453-460.

Cooke, H.J., Schmidtke, J. and Gosden, J.R. (1982) Characterisation of a human Y chromosome repeated sequence and related sequences in higher primates. *Chromosoma* **87**, 491-502.

Corneo, G. (1978) Satellite DNAs in eukaryotes: a major non-adaptive mechanism of speciation which originated with sexual reproduction? *Experientia* **34**, 1141-1142.

Craig, I., Levy, E., and Fraser, N. (1987) The Mammalian Y chromosome: molecular search for the sex-determining gene - summary and perspectives. *Development* **101** (Suppl.), 185-190.

Craig, I.W. and Tolley, E. (1986) Steroid sulphatase and the conservation of the mammalian X chromosome. *Trends in Genetics* **2**, 201-211.

Crampton, J.M., Davies, K.E. and Knapp, T.F. (1981) The occurrence of families of repetitive sequences in a library of cloned cDNA from human lymphocytes. *Nucleic Acids Research* **9**, 3821-3834.

Daiger, S.P., Wildin, R.S. and Su, T.-S. (1982) Sequences on the human Y chromosome homologous to the autosomal gene for argininosuccinate synthetase. *Nature* **298**, 682-684.

d'Ambrosi, E. and Furano, A.V. (1987) DNA synthesis arrest sites at the right terminus of rat long interspersed repeated (LINE or LIRn) DNA family members. *Nucleic Acids Research* **15**, 3155-3175.

Davidson, E.H and Britten, R.J. (1979) Regulation of gene expression: possible role of repetitive sequences. *Science* **20**, 1053-1059.

Davidson, E.H., Hough, B.R., Amanson, C.S. and Britten, R.J. (1973) General interspersion of repetitive with non-repetitive sequence elements in the DNA of *Xenopus*. *Journal of Molecular Biology* **77**, 1-23.

Deeming, D.C. and Ferguson, M.W.J. (1988) Environmental regulation of sex determination in reptiles. *Philosophical Transactions of the Royal Society, London* **B322**, 19-39.

Deininger, P.L. and Schmid, C.W. (1979) A study of the evolution of repeated DNA sequences in primates and the existence of a new class of repetitive sequences in primates. *Journal of Molecular Biology* **127**, 437-460.

Del Villano, B.C. and Lerner, R.A. (1976) Relationship between the oncornavirus gene product gp70 and major protein secretion of the mouse genital tract. *Nature* **259**, 497-499.

de la Chapelle, A. (1987) The Y-chromosome and autosomal testis-determining genes. *Development* **101** (Suppl.), 33-38.

Doolittle, W.F. and Sapienza, C. (1980) Selfish genes, the phenotype paradigm and genome evolution. *Nature* **284**, 601-603.

Dover, G. (1986) Molecular drive in multigene families: how biological novelties arise, spread and are assimilated. *Trends in Genetics* **2**, 159-165.

Eicher, E.M. (1982) Primary Sex Determining Genes in Mice. In *Prospects for Sexing Mammalian Sperm* , Eds. R.A. Amann and G.E. Seidel, Jr. (Colorado Assoc. Uni. Press) pp. 121-132.

Eicher, E.M., Hutchison, K.W., Phillips, S.J., Tucker, P.K. and Lee, B.K. (1989) A repeated segment on the mouse Y chromosome comprised of retroviral-related, Y-enriched and Y-specific sequences. *Genetics* **122**, 181-192.

Eicher, E.M., Phillips, S.J. and Washburn, L.L. (1983) The use of molecular probes and chromosomal rearrangements to partition the mouse Y chromosome into functional regions. In *Recombinant DNA and Medical Genetics*, Eds. Messer, A. and Porter, I.H. (Academic Press, New York) pp. 57-71.

Eicher, E.M. and Washburn,L.L. (1983) Inherited sex reversal in mice: Identification of a new primary sex-determining gene. *Journal of Experimental Zoology* **228**, 297-304.

Eicher, E.M. and Washburn, L.L. (1986) Genetic control of primary sex determination in mice. *Annual Review of Genetics* **20**, 148-149.

Eicher, E.M., Washburn, L.L., Whitney, J.B. and Morrow, K.E. (1982) *Mus poschiavinus* Y chromosome in the C57BL/6 murine genome causes sex reversal. *Science* **217**, 535-537.

Eichwald, E.J. and Silmser, C.R. (1955) (Untitled communication) *Transplantation Bulletin* **2**, 148 -149.

Eisenberg, J.E. (1981) *The Mammalian Radiations.* (The Athlone Press, London).

Ellis, N. and Goodfellow, P.N. (1989) The mammalian pseudoautosomal region. *Trends in Genetics* **5**, 406-410.

Epplen, J.T. (1989) On simple repeated GATA/GACA sequences in animal genomes: a critical reappraisal. *Journal of Heredity* **79**, 409-417.

Epplen, J.T., Cellini, A., Romero, S. and Ohno, S. (1983) An attempt to approach the molecular mechanisms of primary sex determination: W- and Y-chromosomal conserved simple repetitive sequences and their differential expression in mRNA. *Journal of Experimental Zoology* **228**, 305-312.

Epplen, J.T., McCarrey, J.R., Sutou, S. and Ohno, S. (1982) Base sequence of a cloned snake W-chromosome DNA fragment and identification of a male-specific putative mRNA in the mouse. *Proceedings of the National Academy of Sciences, USA* **79**, 3798-3802.

Epplen, J.T. and McLaren, A. (1989) GATA-GACA simple repeated DNA as a means to dissect the sex-reversed (*Sxr*) mutation of the mouse. In *Evolutionary Mechanisms in Sex Determination*, Ed Wachtel, S.S. (CRC Press, Boca Raton, Florida) pp. 171-179.

Epplen, J.T., Studer, R. and McLaren, A. (1988) Heterogeneity in the *Sxr* (sex-reversal) locus of the mouse as revealed by synthetic GATA/GACA probes. *Genetical Research* **51**, 239-246.

Erickson, R.P., Durbin, E.J. and Tres, L.L. (1987) Sex determination in mice: Y and chromosome 17 interactions. *Development* **101** (Suppl.), 25-32.

Erickson, R.P. and Goodfellow, P.N. (1984) Sharing outside of pairing. *Nature* **311**, 106-107.

Erickson, R.P., Lewis, S.E., and Butley, M. (1981) Is haploid gene expression possible for sperm antigens? *Journal of Reproductive Immunology* **3**, 195-217.

Evans, H.J., Buckland, R.A. and Sumner, A.T. (1973) Chromosome homology and heterochromatin in goat, sheep and ox studied by banding techniques. *Chromosoma* **42**, 383-402.

Forejt, J. (1981) Hybrid sterility gene located in the T/t-H-2 supergene on chromosome 17. In *Current Trends in Histocompatibility, vol. 1.* Ed. R.A. Reisfeld and S. Ferrone (Plenum Press, New York) pp. 103-131.

Fredga, K. (1988) Aberrant chromosomal sex-determining mechanisms in mammals with special reference to species with XY females. *Philosophical Transactions of the Royal Society, London* **B322**, 83-95.

Freytag, S.O., Bock, H.O., Beaudet, A.L. and O'Brien, W.E. (1984) Molecular structures of human argininosuccinate synthetase pseudogenes. *Journal of Biological Chemistry* **259**, 3160-3166.

Frommer, M., Prosser, J. and Vincent, P.C. (1984) Human satellite I sequences include a male specific 2.47 kb tandemly repeated unit containing one *Alu* family member per repeat. *Nucleic Acids Research* **12**, 2887-2900.

Geldwerth, D., Bishop, C., Guellaen, G., Koenig, M., Vergnaud, G., Mandel, J. and Weissenbach, J. (1985) Extensive DNA sequence homologies between the human Y and the long arm of the X chromosome. *The EMBO Journal* **4**, 1739-1743.

German, J. (1988) Gonodal dimorphism explained as a dosage effect of a locus on the sex chromosomes, the Gonad-Differentiation Locus (GDL). *American Journal of Human Genetics* **42**, 414-421.

Gillies, C.B. (1974) The nature and extent of synaptonemal complex formation in haploid barley. *Chromosoma* **48**, 441-453.

Goodfellow, P., Banting, G., Sheer, D., Ropers, H.H., Caine, A., Ferguson-Smith, M.A., Povey, S. and Voss, R. (1983) Genetic evidence that a Y-linked gene in man is homologous to a gene on the X chromosome. *Nature* **302**, 346-349.

Goodfellow, P.J., Darling, S.M., Thomas, N.S. and Goodfellow, P.N. (1986) A pseudoautosomal gene in man. *Science* **234**, 740-743.

Goodfellow, P., Darling, S. and Wolfe, J. (1985), The human Y chromosome. *Journal of Medical Genetics* **22**, 329-344.

Goodfellow, P.J., Pritchard, C., Tippett, P. and Goodfellow, P.N. (1987) Recombination between the X and Y chromosomes: implications for the relationship between *MIC2, XG* and *YG*. *Annales of Human Genetics* **51**, 161-167.

Goodman, M., Romero-Herrera, A.E., Dene, H., Czelusnick, J. and Tashian, R.E. (1982) Amino acid sequence evidence on the phylogeny of primates and other eutherians. In *Macromolecular sequences in systematic and evolutionary biology.* Ed. Goodman, M. (Plenum Press, New York) pp. 115-191.

Gordon, J.W. and Ruddle, F.H. (1981) Mammalian gonadal determination and gametogenesis. *Science* **211**, 1265-1271.

Gosden, J.R., Gosden, C.M., Christie, S., Cooke, H.J., Morsman, J.M. and Rodeck, C.H. (1984) The use of cloned Y chromosome-specific DNA probes for fetal sex determination in first trimester prenatal diagnosis. *Human Genetetics* **66**, 347-351.

Graves, J.A.M. (1987) The evolution of mammalian sex chromosomes and dosage compensation: clues from marsupials and monotremes.*Trends in Genetics* **3**, 252-256.

Grimaldi, G., Skowronski, J. and Singer, M. (1984) Defining the beginning and end of *Kpn* I family segments. *EMBO Journal* **3**, 1753-1759.

Grunstein, M. and Hogness, D. (1975) Colony hybridization: a method for the isolation of cloned DNAs that contain a specific gene. *Proceedings of the National Academy of Sciences, USA* **72**, 3961-3965.

Guellaen, G., Casanova, M., Bishop, C., Geldwerth, D., Andre, G., Fellous, M. and Weissenbach, J. (1984) Human XX males with Y single-copy DNA fragments. *Nature* **307**, 172-173.

Hamada, H., Petrino, M.G. and Kaunga, T. (1982) A novel repeated element with Z-DNA-forming potential is widely found in evolutionarily diverse eukaryotic genomes. *Proceedings of the National Academy of Sciences, USA* **79**, 6465-6469.

Hammerberg, C. (1982) The influence of T^{or1} upon male fertility in t-bearing mice. *Genetical Research* **37**, 71-77.

Hanahan, D. (1983) Studies on transformation of *Escherichia coli.* with plasmids. *Journal of Molecular Biology* **166**, 557-580.

Haynes, S.R. and Jelinek, W.R. (1981) Low molecular weight RNAs transcribed *in vitro* by RNA polymerase III from *Alu*-type dispersed repeats in Chinese hamster DNA are also found *in vivo*. *Proceedings of the National Academy of Sciences, USA* **78**, 6130-6134.

Heilig, R., Hanauer, A., Grzeschik, K.H., Hors-Cayla, M.C. and Mandel, J. (1984) Actin-like sequences are present on human X and Y chromosomes. *EMBO Journal* **3**, 1803-1807.

Henikoff, S. (1984) Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. *Gene* **28**, 351-359.

Hentschel, C.C. and Birnstiel, M.L. (1981) The organization and expression of histone gene families. *Cell* **25**, 301-313.

Herr, C., Matthaei, K.I., Holt, N. and Reed, K.C. (1989b) Field implementation of a rapid Y-chromosome-detecting bovine embryo sexing assay. *Australian and New Zealand Society for Cell Biology, 8th Annual Conference*, 51a.

Herr, C., Matthaei, K.I. and Reed, K.C. (1989a) Accuracy of a rapid Y-chromosome-detecting bovine embryo sexing assay. *Australian and New Zealand Society for Cell Biology, 8th Annual Conference*, 50a.

Heslop, B.F., Bradley, M.P. and M.A. Baird, M.A. (1989) A proposed growth regulatory function for the serologically detectable sex-specific antigen H-Ys. *Human Genetics* **81**, 99-104.

Hollis, G.F., Heiter, P.A., McBride, O.W., Swan, D. and Leder, P. (1982) Processed genes: a dispersed human immunoglobulin gene bearing evidence of RNA-type processing. *Nature* **296**, 321-325.

Houck, C.M., Rinehart, F.P. and Schmid, C.W. (1979) A ubiquitous family of repeated DNA sequences in the human genome. *Journal of Molecular Biology* **132**, 289-306.

Ish-Horowicz, D. and Burke, J.F. (1981) Rapid and efficient cosmid cloning. *Nucleic Acids Research* **9**, 2989-2998.

Jackson, J.A. and Fink, G.R. (1981) Gene conversion between duplicated genetic elements in yeast. *Nature* **292**, 306-311.

Jacob, F. (1982) *The Logic of Life.* (Panthean Books, New York).

Jacq, C., Miller, J.R. and Brownlee, G.G. (1977) A pseudogene structure in 5S DNA of *Xenopus laevis*. *Cell* **12**, 109-120.

Jelinek, U.R., Toomey, T.P., Leinward, L., Duncan, C.H., Biro, P.A., Choudary, P.V., Weissman, S.M., Rubin, C.M., Houck, C.M., Deininger, P.L. and Schmid, C.W. (1979) Ubiquitous interspersed repeated sequences in mammalian genomes. *Proceedings of the National Academy of Sciences, USA* **77**, 1398-1402.

Johnson, D.A., Gautsch, J.W., Sportsman, J.R. and Elder, J.H. (1984) Improved technique utilising nonfat dry milk for analysis of proteins and nucleic acids transferred to nitrocellulose. *Gene Analylitical Techniques* **1**, 3-8.

Johnston, R.N., Beverley, S.M. and Schimke, R.T. (1983) Rapid spontaneous dihydrofolate reductase gene amplification shown by fluorescence-activated cell sorting. *Proceedings of the National Academy of Sciences, USA* **80**, 3711-3715.

Jost, A., Vigier, B., Prepin, J. and Perchellet, J.P. (1973) Studies on sex differentiation in mammals. *Recent Progress in Hormone Research* **29**, 1-41.

Kilpatrick, M.W., Klysik, J., Singleton, C.K., Zarling, D.A., Jovin, T.M., Hanau, L.H., Erlanger, B.F. and Wells, R.D. (1984) Intervening sequences in human fetal globin genes adopt left-handed Z helices. *Journal of Biological Chemistry* **259**, 7268-7274.

Keitges, E., Rivest, M., Siniscalco, M. and Gartler, S.M. (1985) X-linkage of steroid sulphatase in the mouse is evidence for a functional Y-linked allele. *Nature* **315**, 226-227.

Kmiec, E.B., Angelides, K.J. and Holloman, W.K. (1985) Left-handed DNA and the synaptic pairing reaction promoted by *Ustilago* Rec1 protein. *Cell* **40**, 139-145.

Kobori, J.A., Strauss, E., Minard, K. and Hood, L. (1986) Molecular analysis of the hotspot of recombination in the murine major histocompatibility complex. *Science* **234**, 173-179.

Koenig, M., Camerino, G., Heilig, R. and Mandel, J. (1984) A DNA fragment from the human X chromosome short arm which detects a partially homologous sequence on the Y chromosome long arm. *Nucleic Acids Research* **12**, 4097-4109.

Koenig, M., Moisan, J.P., Heilig, R. and Mandel, J.L. (1985) Homologies between X and Y chromosomes detected by DNA probes: localisation and evolution. *Nucleic Acids Research* **13**, 5485-5501.

Kole, L.B., Haynes, S.R. and Jelinek, W.R. (1983) Discrete and heterogeneous high molecular weight RNAs complementary to a long dispersed repeat family (a possible transposon) of human DNA. *Journal of Molecular Biology* **165**, 257-286.

Koopman, P., Gubbay, J., Collignon, J. and Lovell-Badge, R. (1989) *Zfy* gene expression patterns are not compatible with a primary role in mouse sex determination. *Nature* **342**, 940-942.

Korn, L.J. and Bogenhagen, D.F. (1982) Organization and transcription of *Xenopus* 5S ribosomal RNA genes. In *The Cell Nucleus,* **12**, Eds Busch, H. and Rothblum, L. (Academic Press, Inc., New York) pp. 1-27.

Kramerov, D.A., Kelak, I.V., Samarina, O. and Ryskov, A.P. (1982) The sequences homologous to major interspersed repeats B1 and B2 of mouse genome are present in mRNA and small cytoplasmic poly(A)+ RNA. *Nucleic Acids Research* **10**, 7477-7491.

Krayev, A.S., Kramerov, D.A., Skryabin, K.G., Ryskov, A.P., Bayev, A.A. and Georgiev, G.P. (1980) The nucleotide sequence of the ubiquitous repetitive DNA sequence B1 complementary to the most abundant class of mouse fold-back RNA. *Nucleic Acids Research* **8**, 1201-1215.

Krco, C.J. and Goldberg, E.H. (1976) H-Y (male) antigen: detection on eight-cell mouse embryos. *Science* **193**, 1134-1135.

Kunkel, L.M. and Smith, K.D. (1982) Evolution of human Y-chromosome DNA. *Chromosoma* **86**, 209-228.

Kunkel, L.M., Smith, K.D. and Boyer, S.H. (1976) Human Y-chromosome-specific reiterated DNA. *Science* **191**, 1189-1190.

Kunkel, L.M., Smith, K.D. and Boyer, S.H. (1979) Organization and heterogeneity of sequences within a repeating unit of human Y chromosome deoxyribonucleic acid. *Biochemistry* **18**, 3343-3353.

Kunkel, L.M., Smith, K.D., Boyer, S.H., Borgaonkar, D.S., Wachtel, S.S., Miller, O.J., Breg, W.R., Jones, H.W. and Rory, J.M. (1977) Analysis of human Y-chromosome-specific reiterated DNA in chromosome variants. *Proceedings of the National Academy of Sciences, USA* **74**, 1245-1249.

Kurnit, D.M., Brown, F.L. and Maio, J.J. (1978) Mammalian repetitive DNA sequences in a stable Robertsonian system. *Cytogenetics and Cell Genetics* **21**, 145-167.

Kurnit, D.M., Shafit, B.R. and Maio, J.J. (1973) Multiple satellite deoxyribonucleic acids in the calf and their relation to the sex chromosomes. *Journal of Molecular Biology* **81**, 273-284.

Lamar, E.E. and Palmer, E. (1984) Y-encoded, species-specific DNA in mice: evidence that the Y chromosome exists in two polymorphic forms in inbred strains. *Cell* **37**, 171-177.

Lau, Y-F. (1987) Localisation of a gene for the male enhanced antigen on human and mouse chromosomes. In *Genetic Markers of Sex Differentiation,* Eds. Haseltine, F.P., McClure, M.E. and Goldberg, E.H. (Plenum Press, New York) pp. 161-167.

Lau, Y-F. (1989) Are male-enhanced antigen and serological H-Y antigen the same? In *Evolutionary mechanisms in sex determination,* Ed. Wachtel, S.S. (CRC Press, Boca Ratan, Florida) pp. 151-157.

Lau, Y.F., Huang, J.C., Dozy, A.H. and Kan, Y.W. (1984) A rapid screening test for antenatal sex determination. *The Lancet* **1**, 14-16.

Lee, M.G., Lewis, S.A., Wilde, C.D. and Cowan, N.J. (1983) Evolutionary history of a multigene family: An expressed human β-tubulin gene and three processed pseudogenes. *Cell* **33**, 477-487.

Leonard, M., Kirszenbaum, M., Cotinot, C., Chesne, P., Heyman, Y., Stinnakre, M.G., Bishop, C., Delouis, C., Vaiman, M. and Fellous, M. (1987) Sexing bovine embryos using a Y chromosome specific DNA probe. *Theriogenology* **27**, 248.

Lerner, R.A., Wilson, C.B., Del Villano, B.C., McConahey, P.J. and Dixon, F.J. (1976) Endogenous oncoviral gene expression in adult and fetal mice: quantitative histologic and physiologic studies of the major viral glycoprotein, gp70. *Journal of Experimental Medicine* **143**, 151-166.

Leroy, P., Seboun, E., Mattei, M.G., Fellous, M. and Bishop, C.E. (1987) Testis-specific transcripts detected by a human Y-DNA-derived probe. *Development* **101** (Suppl.), 177-183.

Levy, E.R. and Burgoyne, P.S. (1986) The fate of XO germ cells in the testes of XO/XY and XO/XY/XYY mouse mosaics: evidence for a spermatogenesis gene on the mouse Y chromosome. *Cytogenetics and Cell Genetics* **42**, 208-213.

Lewin, B. (1974) *Gene expression. 2. Eukaryotic chromosomes.* (Wiley-Interscience Publication, J.Wiley and Sons, New York).

Lewin, B. (1975) Units of transcription and translation: sequence components of heterogeneous nuclear RNA and messenger RNA. *Cell* **4**, 77-93.

Little, P.F.R. (1982) Globin pseudogenes. *Cell* **28**, 683-684.

Lord, E.A. (1989) Repetitive DNA from the ovine Y chromosome. Ph.D. Thesis, Australian National University.

Lucchesi, J.C. (1978) Gene dosage and the evolution of sex chromosomes. *Science* **202**, 711-716.

Lyon, M.F. (1972), X-chromosome inactivation and development patterns in mammals. *Biological Reviews* **47**, 1-35.

Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) *Molecular cloning: A Laboratory Manual.* (Cold Spring Harbor Lab, NY).

Manley, J.L. and Colozzo, M.T. (1982) Synthesis *in vitro* of an exceptionally long RNA transcript promoted by an *Alu* I sequence. *Nature* **300**, 376-379.

Manor, H., Sridhara, R.B. and Martin, R.G. (1988) Abundance and degree of dispersion of genomic d(GA)n.d(TC)n sequences. *Journal of Molecular Evolution* **27**, 96-101.

Manuelidis, L. (1978) Complex and simple sequences in human repeated DNAs. *Chromosoma* **66**, 1-21.

Mardon, G., Mosher, R., Disteche, C.M., Nishioka, Y., McLaren, A. and Page, D.C. (1989) Duplication, deletion and polymorphism in the sex-determining region of the mouse Y chromosome. *Science* **243**, 78-80.

Mardon, G. and Page, D.C. (1989) The sex-determining region of the mouse Y chromosome encodes a protein with a highly acidic domain and 13 zinc fingers. *Cell* **56**, 765-770.

Maroteaux, L., Heilig, R., Dupret, D. and Mandel, J.L. (1983) Repetitive satellite-sequences are present within or upstream from 3 avian protein-coding genes. *Nucleic Acids Research* **11**, 1227-1243.

Matthaei, K.I. and Reed, K.C. (1986) Chromosomal assignment in somatic hybrids by *in situ* hybridization with tritium-labelled Riboprobe RNA probes. *Promega Notes*, **5**, 5.

Matthaei, K.I., Reed, K.C., Nicholls, T.J. and Stephens, I.B. (1990) A practical application of DNA probes in quality assurance of meat products - The sex test. *Australian Society for Animal Production, Proceedings*, (in press).

Matthews, M.E. (1985) Attempted cloning of bovine Y-chromosome-specific DNA. B.Sc.Hons. Thesis, Australian National University.

Matthews, M.E., Matthaei, K.I., Herr, C. and Reed, K.C. (1987) Sex determination of pre-implantation livestock embryos. *Australian Society for Reproductive Biology, Proceedings,* 4.

McCarrey, J.R. and Abbott, U.K. (1979) Mechanisms of genetic sex determination, gonadal sex differentiation and germ-cell development in animals. *Advances in Genetics* **20**, 217-290.

McGraw, R.A., Jacobson, R.J. and Akamatsu, M. (1988) A male-specific repeated DNA sequence in the domestic pig. *Nucleic Acids Research* **16**, 10389.

McKay, R.D.G. Bobrow, M. and Cooke, H.J. (1978) The identification of a repeated DNA sequence involved in the karyotype polymorphism of the human Y chromosome. *Cytogenetics and Cell Genetics* **21**, 19-32.

McKusick, V.A. (1983) *Mendelian Inheritance in Man.* (John Hopkins University Press, Baltimore).

McLaren, A., Simpson,E., Tumonari, K., Chandler, P. and Hogg, H. (1984) Male sexual differentiation in mice lacking H-Y antigen. *Nature* **312**, 552-555.

McLaren, A. (1987) Testis determination and the H-Y hypothesis. *Current Topics in Developmental Biology* **23**, 163-183.

McLaren, A., (1988) Sex determination in mammals. *Trends in Genetics* **4**, 153-157.

Meunier-Rotival, M. and Bernardi, G. (1984) The *Bam* repeats of the mouse genome belong in several superfamilies the largest of which is over 9 kb in size. *Nucleic Acids Research* **12**, 1593-1608.

Miklos, G.L.G. and John, B. (1979) Heterochromatin and satellite DNA in man: properties and prospects. *American Journal of Human Genetics* **31**, 264-280.

Miklos,G.L.G., Matthaei, K.I. and Reed, K.C. (1989) Occurrence of the $(GATA)_n$ sequences in vertebrate and invertebrate genomes. *Chromosoma* **98**, 194-200.

Miller, J., McLachlan, A.D. and Klug, A. (1985) Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus* oocytes. *EMBO Journal* **4**, 1609-1614.

Miller, O.J., Drayna, D. and Goodfellow, P. (1984) Report on the committee on the genetic constitution of the X and Y chromosomes. Human Gene Mapping 7: Seventh International Workshop on Human Gene Mapping. *Cytogenetics and Cell Genetics* **37,** 176-204.

Mittwoch, U. (1986) Males, females and hermaphrodites. *Annales of Human Genetics* **50**, 103-121.

Miyata, T., Hayashida, H., Kuma, K. and Yasanaga, T. (1987) Male-driven molecular evolution demonstrated by different rates of silent substitutions between autosome- and sex chromosome-linked genes. *Proceedings of the Japanese Academy* **63**, 327-331.

Mizusawa, S., Nishimura, S. and Seela, F. (1986) Improvement of the dideoxy chain termination method of DNA sequencing by use of deoxy-7-deazaguanosine triphosphate in place of dGTP. *Nucleic Acids Research* **14**, 1319-1324.

Mondello, C., Ropers, H.H., Craig, I.W., Tolley, E. and Goodfellow, P.N. (1987) Physical mapping of genes and sequences at the end of the human X chromosome short arm. *Annales of Human Genetics* **51**, 137-143.

Moos, M. and Gallwitz, D. (1983) Structure of two human β-actin-related processed genes, one of which is located next to a simple repetitive sequence. *EMBO Journal* **2**, 757-761.

Morton, N.E. (1957) Further scoring types in sequential lineage tests, with a critical review of autosomal and partial sex linkage in man. *American. Journal of Human Genetics* **9**, 55-75.

Moses, M.J. (1968) Synaptonemal complex. *Annual Review of Genetics* **2**, 363-412.

Muller, U. (1987) Mapping of testis-determining locus on Yp by the molecular genetic analysis of XX males and XY females. *Development* **101** (Suppl.), 51-58.

Murray, N.E., Brammer, W.J. and Murray, K. (1977) Lambdoid phages that simplify the recovery of *in vitro* recombinants. *Molecular and General Genetics* **150**, 53-61.

Nagamine, C.M., Chan, K., Kozak, C.A. and Lau, Y.F. (1989) Chromosome mapping and expression of a putative testis-determining gene in mouse. *Science* **243**, 80-83.

Nakazato, H. and Edmonds, M. (1974) Purification of messenger RNA and heterogeneous nuclear RNA containing poly(A) sequences. *Methods in Enzymology* **29**, 431-443.

Needleman, S.B. and Wunsch, C.D. (1970) A general method applicable to the search for similarities in the amino acid sequence of two proteins. *Journal of Molecular Biology* **48**, 443-453.

Nishioka, Y. and Lamothe, E. (1986) Isolation and characterization of a mouse Y chromosomal repetitive sequence. *Genetics* **113**, 417-432.

Nishioka, Y. and Leder, P. (1980) Organization and complete sequence of identical embryonic and plasmacytoma κ V-region genes. *Journal of Biological Chemistry* **255**, 3691-3694.

Nomiyama, H., Obara, K., Jihno, Y., Matsuda, I., Shimada, K. and Miyata, T. (1986) Amplification of human argininosuccinate synthetase pseudogenes. *Journal of Molecular Biology* **192**, 221.

Nordheim, A. and Rich, A. (1983) The sequence (dC-dA)n.(dG-dT)n forms left-handed Z-DNA in negatively supercoiled plasmids. *Proceedings of the National Academy of Sciences, USA* **80**, 1821-1825.

Novacek, M.J. (1982) Information for molecular studies from anatomical and fossil evidence on higher eutherian phylogeny. In *Macromolecular sequences in systematic and evolutionary biology,* Ed. Goodman, M. (Plenum Press, New York) pp. 3-41.

O, W.-S., Short,R.V., Renfree, M.B. and Shaw, G. (1988) Primary genetic control of somatic sexual differentiation in a mammal. *Nature* **331**, 716-717.

Ohno, S. (1967) *Sex chromosomes and sex-linked genes.* (Springer-Verlag, Berlin).

Orgel, L.E. and Crick, F.H.C. (1980) Selfish DNA: the ultimate parasite. *Nature* **284**, 604-607.

Page, D.C. (1988) Is *ZFY* the sex-determining gene on the human Y chromosome? *Philosophical Transactions of the Royal Society, London* ***B322***, 155-157.

Page, D.C., Harper, M.E., Love, J. and Botstein, D. (1984) Occurrence of a transposition from the X-chromosome long arm to the Y-chromosome short arm during human evolution. *Nature* **311**, 119-123.

Page, D.C., Mosher, R., Simpson, E.M., Fisher, E.M.C., Mardon, G. Pollack, J., McGillivray, B., de la Chapelle, A. and Brown, L.G. (1987) The sex-determining region of the human Y chromosome encodes a finger protein. *Cell* **51**, 1091-1104.

Page, G.S., Smith, S. and Goodman, H.M. (1981) DNA sequence of the rat growth hormone gene: location of the 5' terminus of the growth hormone mRNA and identification of an internal transposon-like element. *Nucleic Acids Research* **9**, 2087-2104.

Palmer, M.S., Sinclair, A.H., Berta, P., Ellis, N.A., Goodfellow, P.N., Abbas, N.E. and Fellous, M. (1989) Genetic evidence that *ZFY* is not the testis-determining factor. *Nature* **342**, 937-939.

Petit, C., de la Chapelle, A., Levilliers, J., Castillo, S., Noel, B. and Weissenbach, J. (1987) An abnormal terminal X-Y interchange accounts for most but not all cases of human XX maleness. *Cell* **49**, 595-602.

Phillips, S.J., Birkenmeier, E.H., Callahan, R. and Eicher, E.M. (1982) Male and female mouse DNAs can be discriminated using retroviral probes. *Nature* **297**, 241-243.

Polani, P.E. (1982) Pairing of X and Y chromosomes, non-inactivation of X-linked genes, and the maleness factor. *Human Genetics* **60**, 207-211.

Popescu, C.P. (1973) L'heterochromatine constitutive dans le caryotype bovin normal et anormal. *Annales Genetics* **16**, 183-188.

O, W.-S., Short,R.V., Renfree, M.B. and Shaw, G. (1988) Primary genetic control of somatic sexual differentiation in a mammal. *Nature* **331**, 716-717.

Ohno, S. (1967) *Sex chromosomes and sex-linked genes.* (Springer-Verlag, Berlin).

Orgel, L.E. and Crick, F.H.C. (1980) Selfish DNA: the ultimate parasite. *Nature* **284**, 604-607.

Page, D.C. (1988) Is *ZFY* the sex-determining gene on the human Y chromosome? *Philosophical Transactions of the Royal Society, London* ***B322***, 155-157.

Page, D.C., Harper, M.E., Love, J. and Botstein, D. (1984) Occurrence of a transposition from the X-chromosome long arm to the Y-chromosome short arm during human evolution. *Nature* **311**, 119-123.

Page, D.C., Mosher, R., Simpson, E.M., Fisher, E.M.C., Mardon, G. Pollack, J., McGillivray, B., de la Chapelle, A. and Brown, L.G. (1987) The sex-determining region of the human Y chromosome encodes a finger protein. *Cell* **51**, 1091-1104.

Page, G.S., Smith, S. and Goodman, H.M. (1981) DNA sequence of the rat growth hormone gene: location of the 5' terminus of the growth hormone mRNA and identification of an internal transposon-like element. *Nucleic Acids Research* **9**, 2087-2104.

Palmer, M.S., Sinclair, A.H., Berta, P., Ellis, N.A., Goodfellow, P.N., Abbas, N.E. and Fellous, M. (1989) Genetic evidence that *ZFY* is not the testis-determining factor. *Nature* **342**, 937-939.

Petit, C., de la Chapelle, A., Levilliers, J., Castillo, S., Noel, B. and Weissenbach, J. (1987) An abnormal terminal X-Y interchange accounts for most but not all cases of human XX maleness. *Cell* **49**, 595-602.

Phillips, S.J., Birkenmeier, E.H., Callahan, R. and Eicher, E.M. (1982) Male and female mouse DNAs can be discriminated using retroviral probes. *Nature* **297**, 241-243.

Polani, P.E. (1982) Pairing of X and Y chromosomes, non-inactivation of X-linked genes, and the maleness factor. *Human Genetics* **60**, 207-211.

Popescu, C.P. (1973) L'heterochromatine constitutive dans le caryotype bovin normal et anormal. *Annales Genetics* **16**, 183-188.

Popescu, C.P., Cotinot, C., Boscher, J. and Kirszenbaum, M. (1988) Chromosomal localization of a bovine male specific probe. *Annales Genetics* **31**, 39-42.

Posakony, J.W., Flytzanis, C.N., Britten, R.J. and Davis, E.H. (1983) Interspersed sequence organization and developmental representation of cloned poly(A) RNAs from sea urchin eggs. *Journal of Molecular Biology* **167**, 361-389.

Pritchard, C.A., Goodfellow, P.J. and Goodfellow, P.N. (1987a) Mapping the limits of the human pseudoautosomal region and a candidate sequence for the male-determining gene. *Nature* **328**, 273-275.

Pritchard, C.A., Goodfellow, P.J. and Goodfellow, P.N. (1987b) Isolation of a sequence which maps close to the human sex determining gene. *Nucleic Acids Research* **15**, 6159-6169.

Putney, S.D., Benkovic, S.J. and Schimmel, P.R. (1981) A DNA fragment with an α-phosphorothiate nucleotide at one end is asymmetrically blocked from digestion by exonuclease III and can be replicated *in vivo*. *Proceedings of the National Academy of Sciences, USA* **78**, 7350-7354.

Raman, R. and Nanda, I. (1982) Identification and patterns of synapsis of the autosomally translocated Y chromosome of the indian mongoose *Herpestes auropunctatus* (Hodgson). *Chromosoma* **87**, 477-489.

Reed, K.C. and Mann, D.A. (1985) Rapid transfer of DNA from agarose gels to nylon membranes. *Nucleic Acids Research* **13**, 7207-7221.

Rice, N. R. (1972) In *Evolution of Genetic Systems.*, Ed. Smith, H.H. (Garden Breed, New York) pp 44.

Rich, A. (1983) Right handed and left handed DNA conformational information in genetic material. *Cold Spring Harbor Symposium on Quantitative Biology* **67**, 1-12.

Richter, J.D., Andersen, D.M., Davidson, E.H. and Smith, L.D. (1984) Interspersed poly(A) RNAs of amphibian oocytes are not translatable. *Journal of Molecular Biology* **173**, 227-241.

Rigby, P.W.J., Dieckmann, M., Rhodes, C. and Berg, P. (1977) Labelling deoxyribonucleotide acid to high specific activity *in vitro* by nick translation with DNA Polymerase I. *Journal of Molecular Biology* **113**, 237-251.

Riggs, C.D. and Taylor, J.H. (1987) Sequence organization and developmentally regulated transcription of a family of repetitive DNA sequences of *Xenopus laevis*. *Nucleic Acids Research* **15**, 9551-9565.

Roberts, J.M., Buck, L.B. and Axel, R. (1983) A structure for amplified DNA. *Cell* **33**, 53-63.

Robertson, W.R.B. (1916) Chromosome studies. I. Taxonomic relationships shown in the chromosome of *Tettigidae* and *Acrididae* : V-shaped chromosomes and their significance in *Acrididae*, *Locustidae*, and the *Gryllidae*: chromosomes and variation. *Journal of Morphology* **27**, 179-331.

Roderick, T.H. and Davisson, M.T. (1981) Linkage maps. In: *Genetic variants and strains of the laboratory mouse,* Ed. Green, M.C. (Gustav Fischer, Stuttgart, New York) pp. 279-282.

Roderick, T.H. and Davisson, M.T. (1984) Linkage map of the mouse (*Mus musculus*). In *Genetic Maps 1984,* Ed. O'Brien, S.J. (Cold Spring Harbor Laboratory, New York) pp. 343-356.

Rogers, J. (1985) The origin and evolution of retroposons. *International Review of Cytology* **93**, 187-279.

Roninson, I.G. (1983) Detection and mapping of homologous, repeated and amplified DNA sequences by renaturation in agarose gels. *Nucleic Acids Research* **11**, 5413-5431.

Rouyer, F., Simmler, M., Johnson, C., Vergnaud, G., Cooke, H.J. and Weissenbach, J. (1986) A gradient of sex-linkage in the pseudoautosomal region of the human sex chromosomes. *Nature* **319**, 291-295.

Rubin, G.M. (1983) Dispersed repetitive DNAs in *Drosophila*. In *Mobile Genetic Elements,* Ed. Shapiro, J.A. (Academic Press, New York) pp. 329-361.

Saiki, R.K., Scharf, S., Faloona, F., Mullis, K.B., Horn, G.T., Erlich, H.A. and Arnheim, N. (1985) Enzymatic amplification of β-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. *Science* **230**, 1350-1354.

Sakamoto, K. and Okada, N. (1985) Rodent Type 2 *Alu* family, rat identifier sequence, rabbit C family and bovine or goat 73-bp repeat may have evolved from tRNA genes. *Journal of Molecular Evolution* **22**, 134-140.

Sanger, F., Nicklen, S. and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences, USA* **74**, 5463-5468.

Scherer, S. and Davis, R.W. (1980) Recombination of dispersed repeated sequences in yeast. *Science* **209**, 1380-1384.

Schimenti, J.C. and Duncan, C.H. (1984) Ruminant globin gene structures suggest an evolutionary role for *Alu*-type repeats. *Nucleic Acids Research* **12**, 1641-1655.

Schmeckpeper, B.J., Scott, A.F. and Smith, K.D. (1984) Transcripts homologous to a long repeated DNA element in the human genome. *Journal of Biological Chemistry* **259**, 1218-1225.

Schmid, C.W. and Deininger, P.L. (1975) Sequence organization of the human genome. *Cell* **6**, 345-358.

Schmid, M., Haaf, T., Solleder, E, Schempp, W., Leipold, M. and Heilbronner, H. (1984), Satellited Y chromosomes: structure, origin and clinical significance. *Human Genetics* **67**, 72-85.

Schnedl, W. and Czaker, R. (1974) Centromeric heterochromatin and comparison of G-banding in cattle, goat and sheep chromosomes. *Cytogenetics and Cell Genetics* **13**, 246-255.

Schwartz, C.C., Saffron, W., Welsh, J., Haas, R., Goldenberg, M. and Cantor, C.R. (1982) New techniques for purifying large DNAs and studying their properties and packaging. *Cold Spring Harbor Symposia on Quantitative Biology* **47**, 189-195.

Scott, W.J. and Holsen, J.F. (1977) Weight Differences in Rat Embryos prior to Sexual Differentiation. *Journal of Embryology and Experimental Morph*ology **40**, 259-263.

Scott, A.F. and Smith, K.D. (1982) Genomic DNA: new approaches to evolutionary problems. In *Macromolecular Sequences in Systematic and Evolutionary Biology*, Ed. Goodman, M. (Plenum Press, New York) pp. 319-356.

Seidel, G.E. (1984) Applications of embryo transfer and related technologies to cattle. *Journal of Dairy Science* **67**, 2786-2796.

Seller, M.J. and Perkins-Cole, K.J. (1987) Sex Differences in Mouse Embryonic Development at Neurulation. *Journal of Reproduction and Fertility* **79**, 159-161.

Shelton, J.A. and Goldberg, E.H. (1984) Male-restricted expression of H-Y antigen on preimplantation mouse embryos. *Transplantation* **37**, 7-8.

Silver, J., Rabson, A., Bryan, T., Willey, R. and Martin, M.A. (1987) Human retroviral sequences on the Y chromosome. *Molecular and Cellular Biology* **7**, 1559-1562.

Simmler, M.C., Rouyer, F., Vergnaud, G., Nystrum-Lahti, M., Ngo, K.Y., de la Chapelle, A. and Weissenbach, J. (1985) Pseudoautosomal DNA sequences in the pairing region of the human sex chromosomes. *Nature* **317**, 692-697.

Simpson, E., Chandler, P., Goulmy, E., Page, D.C., Disteche, C., and Ferguson-Smith, M.A. (1987) Separation of the genetic loci for the H-Y antigen and testis determination on human Y chromosome. *Nature* **326**, 876-878.

Sinclair, A.H., Foster, J.W., Spencer, J.A., Page, D.C., Palmer, M., Goodfellow, P.N. and Graves, J.A.M. (1988) Sequences homologous to *ZFY*, a candidate human sex-determining gene, are autosomal in marsupials. *Nature* **336**, 780-783.

Sinclair, A.H. and Graves, J.A.M. (1988) The sex determining gene in mammals: radical new clues from marsupials. *35th Annual Conference of the Genetics Society of Australia, Proceedings,* 43.

Singer, M.F. (1982a) SINEs and LINEs: highly repeated short and long interspersed sequences in mammalian genomes. *Cell* **28**, 433-434.

Singer, M.F. (1982) Highly repeated sequences in mammalian genomes. *International Review of Cytology* **76**, 67-112.

Singer, M.F. and Skowronski, J. (1985) Making sense out of LINEs: long interspersed repeat sequences in mammalian genomes. *Trends in Biochemical Sciences* **10**, 119-122.

Singh, L. and Jones, K.W. (1982) Sex reversal in the mouse (*Mus musculus*) is caused by a recurrent nonreciprocal crossover involving the X and an aberrant Y chromosome. *Cell* **28**, 205-216.

Singh, L., Matsukuma, S. and Jones, K. (1987) The use of Y-chromosome-specfic repeated DNA sequences in the analysis of testis development in an XX/XY mouse. *Development* **101** (Suppl.), 143-149.

Singh, L., Phillips, C. and Jones, K.W. (1984) The conserved nucleotide sequences of Bkm, which define *Sxr* in the mouse, are transcribed. *Cell* **36**, 111-120.

Singh, L., Purdom, I.F. and Jones, K.W. (1980) Sex chromosome associated satellite DNA: evolution and conservation. *Chromosoma* **79**, 137-157.

Slightom, J.L., Blechl, A.E. and Smithies, O. (1980) Human fetal $^{G}\gamma$-and-$^{A}\gamma$-globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. *Cell* **21**, 627-638.

Smith, G.P. (1974) Unequal crossover and the evolution of multigene families. *Cold Spring Harbor Symposium on Quantitative Biology* **38**, 507-513.

Smith, G.P. (1976) Evolution of repeated DNA sequences by unequal crossover. *Science* **191**, 528-535.

Smith, K.D., Young, K.E., Talbot, C.C. and Schmeckpeper, B.J. (1987) Repeated DNA of the human Y chromosome. *Development* **101** (Suppl.), 77-92.

Soares, M.B., Schon, E., Henderson, A., Karathanasis, S.K., Cate, R., Zeitlin, S., Chirgwin, J. and Efstratiadis, A. (1985) RNA-mediated gene duplication: the rat preproinsulin I gene is a functional retroposon. *Molecular and Cellular Biology* **5**, 2090-2103.

Solari, A.J. (1974) The behaviour of the XY pair in mammals. *International Review of Cytology* **38**, 273-317.

Southern, E.M. (1970) Base sequence and evolution of Guinea-pig α-satellite DNA. *Nature* **227**, 794-798.

Staden R. (1982) An interactive graphics program for comparing and aligning nucleic acid and amino acid sequences. *Nucleic Acids Research* **10**, 2951-2961.

Steinmetz, M., Uematsu, Y. and Lindahl, K.F. (1987) Hotspots of homologous recombination in mammalian genomes. *Trends in Genetics* **3**, 7-10.

Thomas, P.S. (1983) Hybridization of denatured RNA transferred or dotted to nitrocellulose paper. *Methods in Enzymology* **100**, 255-266.

Tiepolo, L. and Zuffardi, O. (1976) Localization of factors controlling spermatogenesis in the non-flourescent portion of the human Y-chromosome long arm. *Human Genetics* **34**, 119-124.

Tronick, S.R., Golub, M.M., Stephenson, J.R. and Aaronson, S.A. (1977) Distribution and expression in mammals of genes related to an endogenous Type C RNA virus of *Odocoileus hemionus*. *Journal of Virology* **23**, 1-9.

Van Arsdell, S.W., Denison, R.A., Bernstein, L.B., Weiner, A.M., Manser, T. and Gesteland, R.F. (1981) Direct repeats flank three small nuclear RNA pseudogenes in the human genome. *Cell* **26**, 11-17.

Vanin, E.F. (1984) Processed pseudogenes: characteristics and evolution. *Biochimica et Biophysica Acta* **782**, 231-241.

Vergnaud, G., Kaplan, L., Weissenbach, J., Dumez, Y., Berger, R., Tiollais, P. and Guellaen, G. (1984) Rapid and early determination of sex using trophoblast biopsy specimens and Y chromosome specific DNA probes. *British Medical Journal* **289**, 73-76.

Vergnaud, G., Page, D.C., Simmler, M-C., Brown, L., Rouyer, F., Noel, B., Botstein, D., de la Chapelle, A. and Weissenbach, J. (1986) A deletion map of the human Y chromosome based on DNA hybridisation. *American Journal of Human Genetics* **38**, 109-124.

Vieira, J. and Messing, J. (1982) The pUC plasmids, an M13mp7-derived system for insertion mutagenesis and sequencing with synthetic universal primers. *Gene* **19**, 259-268.

Vieira, J. and Messing, J. (1987) Production of single-stranded plasmid DNA. *Methods in Enzymology* **153**, 3-11.

Vrati, S., Mann, D.A., and Reed, K.C., (1987) Alkaline northern blots: transfer of RNA from agarose gels to Zeta-Probe membrane in dilute NaOH. *Bio-Rad Mol. Biol. Reports* **1**, 1-4.

Wachtel, S.S., Ohno, S., Koo, G.C., and Boyse, E.A. (1975) Possible role for H-Y antigen in the primary determination of sex. *Nature* **257**, 235-236.

Wang, A.H.-J., Quigley, G.J., Kolpak, F.J., Crawford, J.L., van Boom, J.H., van der Marel, G. and Rich, A. (1979) Molecular structure of a left-handed double helical DNA fragment at atomic resolution. *Nature* **282**, 680-686.

Waring, M. and Britten, R.J. (1966) Nucleotide sequence repetition: a rapidly reassociating fraction of mouse DNA. *Science* **154**, 791-794.

Watanabe, Y., Tsukada, T., Notake, M., Nakanishi, S. and Numa, S. (1982) Structural analysis of repetitive DNA sequences in the bovine corticotrophin-β-lipotropin precursor gene region. *Nucleic Acids Research* **10**, 1459-1469.

Watson, J.D., Hopkins, N.H., Roberts, J.W., Steitz, J.A. and Weiner, A.M. (1987) *Molecular Biology of the Gene, 4th edition* (Benjamin/ Cummings, Menlo Park, California).

Weber, B., Walz, L., Schmid, M. and Schempp, W. (1988) Homeologic aberrations in human and chimpanzee Y chromosomes: inverted and satellited Y chromosomes. *Cytogenetics and Cell Genetics* **47**, 26-28.

Weiner, A.M., Deininger, P.L. and Efstratiadis, A. (1986) Nonviral retroposons: genes, pseudogenes and transposable elements generated by the reverse flow of genetic information. *Annual Review of Biochemistry* **55**, 631-661.

Weissenbach, J., Levilliers, J., Petit, C., Rouyer, F. and Simmler, M. (1987) Normal and abnormal interchanges between the human X and Y chromosomes. *Development* **101** (Suppl.), 67-74.

West, J.D., Gosden, J.R., Angell, R.R., Hastie, N.D., Thatcher, S.S., Glaier, A.F. and Baird, D.T. (1987) Sexing the human pre-embryo by DNA-DNA *in situ* hybridization. *The Lancet* **1**, 1345-1347.

Wiberg, U.H. (1987) Facts and considerations about sex-specific antigens. *Human Genetics* **76**, 207-219.

Williams, T.J. (1986) A technique for sexing mouse embryos by a visual colorimetric assay of the X-linked enzyme, glucose-6-phosphate dehydrogenase. *Theriogenology* **25**, 733-739.

Wilson, A.C., Sarich, V.M. and Maxson, L.R. (1974) The importance of gene rearrangement in evolution: evidence from studies on rates of chromosomal, protein and anatomical evolution. *Proceedings of the National Academy of Sciences, USA* **71**, 3028-3030.

Wolfe, J. Darling, S.M., Erickson, R.P., Craig, I.W., Buckle, V.J., Rigby, P.W.J., Willard, H.F. and Goodfellow, P.N. (1985) Isolation and characterization of an alphoid centromeric repeat family from the human Y chromosome. *Journal of Molecular Biology* **182**, 477-485.

Wolfe, J. Erickson, R.P., Rigby, P.W.J. and Goodfellow, P.N. (1984), Cosmid clones derived from both euchromatic and heterochromatic regions of the human Y chromosome. *EMBO Journal* **3**, 1997-2003.

Wurster, D.H. and Benirschke, K. (1968) Chromosome studies in the superfamily *Bovoidea*. *Chromosoma* **25**, 152-171.

Yanisch-Perron, C., Vieira, J. and Messing, J. (1985) Improved M13 phage cloning vectors and host strains: nucleotide sequences of the M13mp18 and pUC19 vectors. *Gene* **33**, 103-119.

APPENDIX 1: DNA SEQUENCE OF BRY.1

```
        10         20         30         40         50
GATCCGAGACACAGAACAGGCTGCAATTCCAATACACAGAGGTCATGGTG
CTAGGCTCTGTGTCTTGTCCGACGTTAAGGTTATGTGTCTCCAGTACCAC
Sau 3AI

        60         70         80         90        100
GGTGACCCCACGGCCCTTTGGACGTGCAGCTACAAGGGCTTTCTATCCTT
CCACTGGGGTGCCGGGAAACCTGCACGTCGATGTTCCCGAAAGATAGGAA

       110        120        130        140        150
CTAGAACACTGGGCTTTCTTTCCCTGTCCTTGCCCTGAGCATGAGCACCC
GATCTTGTGACCCGAAAGAAAGGGACAGGAACGGGACTCGTACTCGTGGG

       160        170        180        190        200
CTTGCCTTTTTTTCTGAGGTTTCAGAAATGGACCAGCACTGCAGCATAAGC
GAACGGAAAAAAAGACTCCAAAGTCTTTACCTGGTCGTGACGTCGTATTCG

       210        220        230        240        250
ACCTGCTACCTGCATAGTCTTCCAGTTTGAAAAATCACTCTTTGTACTCT
TGGACGATGGACGTATCAGAAGGTCAAACTTTTTAGTGAGAAACATGAGA

       260        270        280        290        300
TTGAAGAAGGCATATATTCGGAGTAAGGACTATAGGATGGATGGATTAGC
AACTTCTTCCGTATATAAGCCTCATTCCTGATATCCTACCTACCTAATCG

TTGATC
AACTAG
Sau 3AI
```

APPENDIX 2: DNA SEQUENCE OF BRY.2

```
        10         20         30         40         50
GGATCCCGCAGTGAATCATCCACACCCTTTCACACTGACAGTCATCACCC
CCTAGGGCGTCACTTAGTAGGTGTGGGAAAGTGTGACTGTCAGTAGTGGG
```

***Bam* HI** **1st region of homology with OY1 begins (70% identity)**

```
        60         70         80         90        100
AAGGAGCCACGTTTTACCAGTGATGTAAGGTAATAGCTCTTTAATGATCA
TTCCTCGGTGCAAAATGGTCACTACATTCCATTATCGAGAAATTACTAGT

       110        120        130        140        150
ACTGTTGTTCAAGCAAGCTAAAGGGTGGTCCAAAAAGGAAATATCAGCTT
TGACAACAAGTTCGTTCGATTTCCCACCAGGTTTTTCCTTTATAGTCGAA

       160        170        180        190        200
GCTGTGGACCAGAGATGGCTTGCACCTGTACCTGCTAGGATAGGAGAAGG
CGACACCTGGTCTCTACCGAACGTGGACATGGACGATCCTATCCTCAACC
```

1st region of homology with OY1 ends

```
       210        220        230        240        250
AAAAGGTGCAAGCTTCTAGGGGCCTCAGCTTGCCTCCCATTAACAGCTTC
TTTTCCACGTTCGAAGATCCCCGGAGTCGAACGGAGGGTAATTGTCGAAG

       260        270        280        290        300
CCTGAAGCTCTCAACTCTCACAGTGTAGAACATGATAACATGATCCTCCA
GGACTTCGAGAGTTGAGAGTGTCACATCTTGTACTATTGTACTAGGAGGT

       310        320        330        340        350
CTCTCCACCAAGTCCCATACCCACCTGTCTCAGTCTCCCTCATTGACCTT
GAGAGGTGGTTCAGGGTATGGGTGGACAGAGTCAGAGGGAGTAACTGGAA

       360        370        380        390        400
CAAGTTGATCGTGTAGCTGAGAAAGTCTTTATCTTGGTCGCTGATGATGA
GTTCAACTAGCACATCGACTCTTTCAGAAATAGAACCAGCGACTACTACT

       410        420        430        440        450
CAGAAACTTGAGGGTGGTTCATAATCTGCTTTAGGTCAAGGAGCCATTTA
GTCTTTGAACTCCCACCAAGTATTAGACGAAATCCAGTTCCTCGGTAAAT

       460        470        480        490        500
GGCCACATACTAGGCCCAAGAAGATCCCAGGAACAACAGTCCTAGTCTGG
CCGGTGTATGATCCGGGTTCTTCTAGGGTCCTTGTTGTCAGGATCAGACC
```

```
       510        520        530        540        550
GGTGGAAGAAGGACACTGGAACAGGCATCGGCAAGGATGCTGGGCACTTC
CCACCTTCTTCCTGTGACCTTGTCCGTAGCCGTTCCTACGACCCGTGAAG

       560        570        580        590        600
TCTTCCCAGCCACTACTGATGTCCTCAGCCCTGCCATACCATCAGACACG
AGAAGGGTCGGTGATGACTACAGGAGTCGGGACGGTATGGTAGTCTGTGC

       610        620        630        640        650
CTCATGGCTCTGGGTTCCTTTTGGTGTTCTGCATCTGAGGTGGTCTGCAT
GAGTACCGAGACCCAAGGAAAACCACAAGACGTAGACTCCACCAGACGTA

       660        670        680        690        700
TCCCCAAGCAGGGACTGTTTCTGATTGTTATTCTTCTCTAGAGTAGTGTA
AGGGGTTCGTCCCTGACAAAGACTAACAATAAGAAGAGATCTCATCACAT

       710        720        730        740        750
TGTGTGTTGTGTGTGTGTATGTGTGACTCCCATATCTCAGTTATTTTCGG
ACACACAACACACACACATACACACTGAGGGTATAGAGTCAATAAAAGCC
(TG).(AC)11

       760        770        780        790        800
GATCATCCACATTTACAACAATAACTGTCCGCTGAGACTGCTCTGGTGTC
CTAGTAGGTGTAAATGTTGTTATTGACAGGCGACTCTGACGAGACCACAG

       810        820        830        840        850
TGTCTCTGGTCATTCCTGACTTAGAGGCCTCGGAGCTGTTCCTCCTCAGA
ACAGAGACCAGTAAGGACTGAATCTCCGGAGCCTCGACAAGGAGGAGTCT

       860        870        880        890        900
TGCAGGCCTGAAGTAGACACACAATGTCATGGCTGCGACACTCACAGGTG
ACGTCCGGACTTCATCTGTGTGTTACAGTACCGACGCTGTGAGTGTCCAC

       910        920        930        940        950
TGGCCCTGCCAGGACTCTATTCAGCTGTGCCTTTGAGCCTGCGAATTGCC
ACCGGGACGGTCCTGAGATAAGTCGACACGGAAACTCGGACGCTTAACGG

       960        970        980        990       1000
AGTACCACACTTCTTCAGGTCACAGCACATATCTGTGCCCTCGCCCCCAC
TCATGGTGTGAAGAAGTCCAGTGTCGTGTATAGACACGGGAGCGGGGGTG
```

```
      1010       1020       1030       1040       1050
CCCTCCTCCTGCTCCTGCCTCTGCACCTCCTTCTCCTCCTCCTAAGTAGA
GGGAGGAGGACGAGGACGGAGACGTGGAGGAAGAGGAGGAGGATTCATCT

      1060       1070       1080       1090       1100
CCCCAAGGAGGCAGCAGAAAGGATACCACTTTGTCCCAGAAGCCAGGGAT
GGGGTTCCTCCGTCGTCTTTCCTATGGTGAAACAGGGTCTTCGGTCCCTA

      1110       1120       1130       1140       1150
GCCCTGGATGGTGGCGCTCCTCTAAGCCAAGTCACAATTCCTCCTCTGAT
CGGGACCTACCACCGCGAGGAGATTCGGTTCAGTGTTAAGGAGGAGACTA

      1160       1170       1180       1190       1200
GGCTCTTCACATGAACCAAACATAGGCCCTGTGGTTTTTCTCATGCTAAG
CCGAGAAGTGTACTTGGTTTGTATCCGGGACACCAAAAAGAGTACGATTC
```

2nd region of homology with OY1 begins (77% identity)

```
      1210       1220       1230       1240       1250
ACGTCAGTTCCACTTGCAGGGCCGGCCAGCGCCTGCAGCACATCTAGCAC
TGCAGTCAAGGTGAACGTCCCGGCCGGTCGCGGACGTCGTGTAGATCGTG

      1260       1270       1280       1290       1300
CAGAAGCTCATTCAGGCCTTTCCTGGCCCTGCTCCTCCAACATCTTCTCC
GTCTTCGTGTAAGTCCGGAAAGGACCGGGACGTGGAGGTTGTAGAAGAGG

      1310       1320       1330       1340       1350
TCCAGCTCCTGGGAGGACCCCTGTTGCTGCTCCTCATCAGCTGAAACCTC
AGGTCGAGGACCCTCCTGGGGACAACGACGAGGAGTAGTCGACTTTGGAG

      1360       1370       1380       1390       1400
TACCTCCTCCATGATGTCTTCCACAAGCAGCTGGAACACCCGCCTGATCC
ATGGAGGAGGTACTACAGAAGGTGTTCGTCGACCTTGTGGGCGGACTAGG

      1410       1420       1430       1440       1450
CCGCCACATCTCCATCCACCAGGGACCCGCCGCATCCTCTGCCACCTCCA
GGCGGTGTAGAGGTAGGTGGTCCCTGGGCGGCGTAGGAGACGGTGGAGGT

      1460       1470       1480       1490       1500
CCCTGAAGAGGGTGGCATCCTTCACCTTGTGTGACAACTGGTGGCTGCTA
GGGACTTCTCCCACCGTAGGAAGTGGAACACACTGTTGACCACCGACGAT
```

```
      1510       1520       1530       1540       1550
GGTCCCAGCCGTGCCATAGCATCACAAGGGCAGGCCACGGGGCCCATCCC
CCAGGGTCGGCACGGTATCGTAGTGTTCCCGTCCGGTGCCCCGGGTAGGG

      1560       1570       1580       1590       1600
CCTGAACACCTGGGCTCACAACTAAGAAGGTCCGGGGTCCCAGGACACTC
GGACTTGTGGACCCGAGTGTTGATTCTTCCAGGCCCCAGGGTCCTGTGAG

      1610       1620       1630       1640       1650
CCACATTCCTATGACACTCTCACGCTTCTCTTGGCCCTGGCCATAACCTC
GGTGTAAGGATACTGTGAGAGTGCGAAGAGAACCGGGACCGGTATTGGAG
```

2nd region of homology with OY1 ends

```
      1660       1670       1680       1690       1700
CGGCTGTTAGAAGCGAAAGGACAGACGTAATACCACCAAAGCAAGTGGTG
GCCGACAATCTTCGCTTTCCTGTCTGCATTATGGTGGTTTCGTTCACCAC

      1710       1720       1730       1740       1750
TCCAATGGAGACCCAGTCTGCAGTAACCTGTGGGGTCCGCTCTTCTCGGG
AGGTTACCTCTGGGTCAGACGTCATTGGACACCCCAGGCGAGAAGAGCCC

      1760       1770       1780       1790       1800
TCGCGGAGTGAGAGGCATGGTGGAGGCAGGGCTGGTGTTTGGGTGCCTGA
AGCGCCTCACTCTCCGTACCACCTCCGTCCCTCCCACAAACCCACGGACT
```

1st copy of 52 nt direct repeat

```
      1810       1820       1830       1840       1850
GATGAAGAAGGGTAGTGACATCCAGATAGCTGACGGCTGGAACTACAACG
CTACTTCTTCCCATCACTGTAGGTCTATCGACTGCCGACCTTGATGTTGC

      1860       1870       1880       1890       1900
CGGCTAGTGTGACGATCCAGCTGCTGCCAATTAGACATGCGAAGAGTGGC
GCCGATCACACTGCTAGGTCGACGACGGTTAATCTGTACGCTTCTCACCT

      1910       1920       1930       1940       1950
GATCTTGAACTCCTAAGTCATCGGATAGGATAGATGATGATTGATTAGCT
CTAGAACTTGAGGATTCAGTAGCCTATCCTATCTACTACTAACTAATCGA

      1960       1970       1980       1990       2000
TGATCCTCTTCGTGCTCCGGCCCACAGATTGGCCATTGGGAGGTGGTACG
ACTAGGAGAAGCACGAGGCCGGGTGTCTAACCGGTAACCCTCCACCATGC
```

131 nt inverted repeat begins 1st copy

2010 2020 2030 2040 2050
TGTTCTTACATGGAGAACAGAACCACTTTGAAAGGGCCTTGATAGGGGTT
ACAAGAATGTACCTCTTGTCTTGGTGAAACTTTCCCGGAACTATCCCCAA

2060 2070 2080 2090 2100
AAATGGGCAGACATGCATGTCTCAGGCACCAGGAAGTCATGGAACTTAAC
TTTACCCGTCTGTACGTACAGAGTCCGTGGTCCTTCAGTACCTTGAATTG
inverted repeat ends nt 2097; 3rd region of homology with OY1 begins

2110 2120 2130 2140 2150
AGGGAGTAGGGGCTGGTGTGTCTGGGGTGGCCTGAGGATGAAGAAAGGGG
TCCCTCATCCCCGACCACACAGACCCCACCGGACTCCTACTTCTTTCCCC
at nt 2074 (62% identity), 2nd copy of long direct repeat starts at nt 2119

2160 2170 2180 2190 2200
TAGTGGACATCCCAGGATAGGCTGAACGGGCCTGGGACATCACCCAAACA
ATCACCTGTAGGGTCCTATCCGACTTGCCCGGACCCTGTAGTGGGTTTGT
(11 mismatches) 2nd copy ends at nt 2185

2210 2220 2230 2240 2250
GGCAGGGCCCTAGGTTGTGAGCCATGCACAGGCTGCTGGCCCAAATCTAG
CCGTCCCGGGATCCAACACTCGGTACGTGTCCGACGACCGGGTTTAGATC

2260 2270 2280 2290 2300
ACCAATGGCCGAAAGGGTAGTGGACGGCATCCTTATGGACATGCCCCTAG
TGGTTACCGGCTTTCCCATCACCTGCCGTAGGAATACCTGTACGGGGATC

2310 2320 2330 2340 2350
ACTGGAGGGCATCTTCCCTGGTTCCTGGGCTCCAGCTCCACCACGTGCCA
TGACCTCCCGTAGAAGGGACCAAGGACCCGAGGTCGAGGTGGTGCACGGT

2360 2370 2380 2390 2400
GATTCTGCTCCATGGAGCTAGCAGGTTTTGCATCTGCCGCTCCCAGGGGC
CTAAGACGAGGTACCTCGATCGTCCAAAACGTAGACGGAGAGGGTCCCCG

2410 2420 2430 2440 2450
TGTCATTGACCTGGATCTGGAAATTCCAACACACGGTTCCTGCTCCGCAA
ACAGTAACTGGACCTAGACCTTTAAGGTTGTGTGCCAAGGACGTGGCGTT

2460 2470 2480 2490 2500
TCCGAAAAACACACTCCTGGGTCCCACAAGACCTTTAATAAATGGGCGTC
AGGCTTTTTGTGTGAGGACCCAGGGTGTTCTGGAAATTATTTACCCGCAG

2510 2520 2530 2540 2550
TCCTGAGTATCAAAAGATACGCAGGGGCACCAAGTGGCTGGATACTTTTC
AGGACTCATAGTTTTCTATGCGTCCCCGTGGTTCACCGACCTATGAAAAG

2560 2570 2580 2590 2600
TCCATGGGACAGAGTACCTGTTGCTGTCAGATGCTATGGGAGTAGGGAAA
AGGTACCCTGTCTCATGGACAACGACAGTCTACGATACCCTCATCCCTTT

2610 2620 2630 2640 2650
ATCCCACCATTTGCAACAGTTGATCTGCATTGTTCGGGACACCAGAACCT
TAGGGAGGTAAACGTTGTCAACTAGACGTAACAAGCCCTGTGGTCTTGGA

2660 2670 2680 2690 2700
TAGGTACAGAGCTCTGTGTCTCTTTATCTCTGCTTCTGGCATAGCACTGT
ATCCATGTCTCGAGACACAGAGAAATAGAGACGAAGACCGTATCGTGACA

2nd copy of

2710 2720 2730 2740 2750
TTTGGGCTATCCTTCTGTGTGTGTACCAGGGCTGGTGTCTATGTAGTTCC
AAACCCGATAGGAAGACACACACATGGTCCCGACCACAGATACATCAAGG

131 nt inverted repeat begins nt 2690 (19 mismatches)

2760 2770 2780 2790 2800
ATCTCTTTAAGTGAATGCTGTTACCCTTTGCCACTGTCTGGACACCAGCA
TAGAGAAATTCACTTACGACAATGGGAAACGGTGACAGACCTCTGGTCGT

3rd region of homology with OY1 ends at nt 2778

2810 2820 2830 2840 2850
CTCATACGAGAAGCTTATCCTTGGTGTGAAGGCAAGCCCTTCTCCTCCTG
GAGTATGCTCTTCGAATAGGAACCACACTTCCGTTCGGGAAGAGGAGGAC

2nd copy of inverted repeat ends

2860 2870 2880 2890 2900
AGTGAGTTTCACTAATGGGAGATCAGACTCTTTTTTAATTTAAATTTATT
TCACTCAAAGTGATTACCCTCTAGTCTGAGAAAAAATTAAATTAAAATAA

2910 2920 2930 2940 2950
TATTTTAAGTAGAGGCTAATTACTTTACAATATTGTATTGGTTTTGCCAT
ATAAAATTCATCTCCGATTAATGAAATGTTATAACATAACCAAAACGGTA

2960 2970 2980 2990 3000
ACATCAACATGAGTCCACCATGGGTGTACACAATCAGACATTTTTAGTCC
TGTAGTTGTACTCAGGTGGTACCCACATGTGTTAGTCTGTAAAAATCAGG

4th region of homology with OY1 begins (64% identity)

3010 3020 3030 3040 3050
ACACATTCAGACAGTGCTTTCTCAAACTTGAAAGCAACAGTGGCGTGGTC
TGTGTAAGTCTGTCACGAAAGAGTTTGAACTTTCGTTGTCACCGCACCAG

3060 3070 3080 3090 3100
ACTGCTGAACCTCAGTTCAGGGTCGGTTATCCTCCTCCTCGCTGGACAGA
TGACGACTTGGAGTCAAGTCCCAGCCAATAGGAGGAGGAGCGACCTGTCT

3110 3120 3130 3140 3150
TGTGAACACTGCAGGTCTGTAGTGTCTTGCACACTTCGACAGAGGAGAAT
ACACTTGTGACGTCCAGACATCACAGAACGTGTGAAGCTGTCTCCTCTTA

3160 3170 3180 3190 3200
GGCATGCTCTGCTGGCATGAGGAACTCTGACTGTCAATGCTACAGTGAAT
CCGTACGAGACGACCGTACTCCTTGAGACTGACAGTTACGATGTCACTTA

3210 3220 3230 3240 3250
GCTAGCAACAACTCCCTGTTAAGTTCCATGACTTCCTGGTGCCTGAGACA
CGATCGTTGTTGAGGGACAATTCAAGGTACTGAAGGACCACGGACTCTGT

3260 3270 3280 3290 3300
TGCATGTCTGCCCATTTACCCTATCAAGCCCTTTCAAAGTGGTTCTGTTC
ACGTACAGACGGGTAAATGGGATAGTTCGGGAAAGTTTCACCAAGACAAG

3310 3320 3330 3340 3350
TCCATGTAAGAACACGTACCAGCCTGCCCAATCGCCGAATCCTGTGGGGC
AGGTACATTCTTGTGCATGGTCGGACGGGTTAGCGGCTTAGGACTCCCCG

3360 3370 3380 3390 3400
CTGTGTGGGGCCGGGAGCAGGTCGAAGAGGATCAAGCTAATCCATCCATG
GACACACCCCGGCCCTCGACCAGCTTCTCCTAGTTCGATTAGGTAGGTAC

BRY.1 homologue 6 mismatches

3410 3420 3430 3440 3450
CCTGATAGTCCGTTACTCCGAATATATGCCTTCTTCAGAAGAGTAGCAAA
GGACTATCAGGCAATGAGGCTTATATACGGAAGAAGTCTTCTCATCGTTT

3460 3470 3480 3490 3500
GAGTGATTTTTCAAACTGGAAGACTATGCAGGTAGCAGGTGCTTATGCTG
CTCACTAAAAAGTTTGACCTTCTGATACGTCCATCGTCCACGAATACGAC

```
      3510       3520       3530       3540       3550
CAGTGGCTGGTCCATTTCTGAAACCTCAGAAAAAAGGCAAGGGGTGCTCA
GTCACCGACCAGGTAAAGACTTTGGAGTCTTTTTTCCGTTCCCCACGAGT

      3560       3570       3580       3590       3600
TGCTCAGGGCAAGGACAGGGAAAGAAAGCCCAGTGTTCTAGAAGGATAGA
ACGAGTCCCGTTCCTGTCCCTTTCTTTCGGGTCACAAGATCTTCCTATCT

      3610       3620       3630       3640       3650
AAGCCCTTGTAGCTGCACGTCCAAAGGGCCGTGGGGTCACCCACCATGAC
TTCGGGAACATCGACGTGCAGGTTTCCCGGCTCCCCAGTGGGTGGTACTG

      3660       3670       3680       3690
CTCTGTGTATTGGAATTGCAGCCTGTTCTGTGTCTCGGATCC
GAGACACATAACCTTAACGTCGGACAAGACACAGAGCCTAGC
```

BRY.1 homologue and OY1 homology ***Bam*** **HI**

APPENDIX 3: DNA SEQUENCE OF BRY.3A

```
        10         20         30         40         50
AGGATCCGAGCTCGTACCGGGATCAGGATACACTCAGATGAATGCGCCGC
TCCTAGGCTCGAGCATGGCCCTAGTCCTATGTGAGTCTACTTACGCGGCG
```

***Bam* HI**

```
        60         70         80         90        100
GAGACAGAAAGTAAAGGAGAGAAAAAGAGGTGATACCCCTTGGTTTACAC
CTCTGTCTTTCATTTCCTCTCTTTTTCTCCACTATGGGGAACCAAATGTG

    110        120        130        140        150
AGAAAGCCAGTAAAGGCCTGACATGGAACTTGCTCTGTTCACAGAGGCCA
TCTTTCGGTCATTTGGCCACTGTACCTTGAACGAGACAAGTGTCTCCGGT

    160        170        180        190        200
CAGGCACCCTCTCAATGGAGCAAAGACGCAAAGGCTTCTCCTGGAGCGAA
GTCCGTGGGAGAGTTACCTCGTTTCTGCGTTTCCGAAGAGGACCTCGCTT

    210        220        230        240        250
GACACAGAGCGCCTTCTCAATTGGGTCTTAGAAGCCTGGGCAAAAAAGTG
CTGTGTCTCGCGGAACTGTTAACCCAGAATCTTCGGACCCGTTTTTTCAC

    260        270        280        290        300
AACTCAGAGAGCCTCTGTGCTCTAGTCCAGGATCTGCTGAGAAACAGAGA
TTGAGTCTCTCGGAGACTCGAGATCAGGTCCTAGACGACTCTTTGTCTCT

    310        320        330        340        350
AGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG
TCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTC
```

homopurine tract: $(GA)_{24}$ $(GGAGA)_6$ $(GA)_{21}$

```
    360        370        380        390        400
GAGAGGAGAGGAGAGGAGAGGAGAGAGAGAGAGAGAGAGAGAGAGAG
CTCTCCTCTCCTCTCCTCTCCTCTCTCTCTCTCTCTCTCTCTCTCTC

    410        420        430        440        450
AGAGAGAGAGAGAGATGACGGACAGTAGGGAGTTATCGTTGGAGTGTTGT
TCTCTCTCTCTCTCTACTGCCTGTCATCCCTCAATAGCAACCTCACAACA

    460        470        480        490        500
GAACAATATGCTCTATCTAGTGAGTGTGGACATATGGGTAGTGAGAGTGT
CTTGTTATACGAGATAGATCTCTCACACCTGTGTGCCCATCACTCTCACA
```

```
      510        520        530        540        550
GATAAATGAGAAGTAAATAGTACAGAGGGTAGATAAATGTCTTTAGCTAG
CTATTTACTCTTCATTTATCATGTCTCCCATCTATTTACAGAAATCGATC

      560        570        580        590        600
AGATTATTAACTAGAGCTTAAGTTGATTTTCTTCAGAGAATGGTGGTCAG
TCTAATAATTGATCTCGAATTCAACTAAAAGTTGTCTCTTACCACCAGTC

      610        620        630        640        650
GGAGAGACACTGTTAATGTCAGAAGAACTGGTGAAAGGCATAATATAGTA
CCTCTCTGTGACAATTACAGTCTTCTTGACCACTTTCCGTATTATATCAT

      660        670        680        690        700
AACAGTAGACAGATAGATTTTGGTTTTCGGCGGCAGATGTTTCGACGTAG
TTGTCATCTGTCTATCTAAAACCAAAAGCCGCCGTCTACAAAGCTGCATC

      710        720        730        740        750
GTCCAGGGAATCCCTCGAGTCTGATCCGCCTTTTGCCTATAGCGTCTTCT
CAGGTCCCTTAGGGAGCTCAGACTAGGCGGAAAACGGATATCGCAGAAGA

      760        770        780        790        800
CCGCACTCGCCTTGTCATGCGTGGGATCTCCTGTGATCGGCTCCTGACCA
GGCGTGAGCGGAACAGTACGCACCCTAGAGGACACTAGCCGAGGACTGGT

      810        820        830        840        850
AACGGGGTTCTCGTTCTCCATGTGAGAACACTATCCAGGCTGCCGGGTCG
TTGCCCCAAGAGCAAGAGGTACACTCTTGTGATAGGTCCGACGGCCCAGC
```

homology with OY1 begins (70% identity)

```
      860        870        880        890        900
CCAAGGTGACCACCGGGTGGCAGAGCAGCAAGAGGATCAAGCTAAGCCAT
GGTTCCACTGGTGGCCCACCGTCTCGTCGTTCTCCTAGTTCGATTCGGTA

      910        920        930        940        950
CATCTGTAGCCTGATACCAATATATGTGTTCTCGAAGTATACAATAGTGA
GTAGACATCGGACTATGGTTATATACACAAGAGCTTCATATGTTATCACT

      960        970        980        990       1000
TGTTCAACTCCTCTAGTGTCACCGTGCAGATCCCCTACAACAAAATCAGA
ACAAGTTGAGGAGATCACAGTGGCACGTCTAGGGGATGTTGTTTTAGTCT
```

```
      1010        1020        1030        1040        1050
GAATCTGTTCCAGAGAGGACTGCCAGAATCCTCTCTCAGTTGCAGTCGCA
CTTAGACAAGGTCTCTCCTGACGGTCTTAGGAGAGAGTCAACGTCAGCGT

      1060        1070        1080        1090        1100
TATCTTGGCCCCTGTAGACCCCACAGCCAGGGGCATGACCCTAGAGGAAA
ATAGAACCGGGGACATCTGGGGTGTCGGTCCCCGTACTGGGATCTCCTTT

      1110        1120        1130        1140        1150
TAAGGAATTCACGTCAGGACTTTCAGCACAGCCATGTGTGTCTTTGCAGG
ATTCCTTAAGTGCAGTCCTGAAAGTCGTGTCGGTACACACAGAAACGTCC

      1160        1170        1180        1190        1200
ACGGCTATTTGCCTAAACCCTGGACCCTGAAAACCAGCAGGCTTCAAATT
TGCCGATAAACGGATTTGGGACCTGGGACTTTTGGTCGTCCGAAGTTTAA

      1210        1220        1230        1240        1250
ACCCCAGGGGCTTCCTGTCTCCACTCTCTGAGTTCCTTGGACATGCATAA
TGGGGTCCCCTAAGGACAGAGGTGAGAGACTCAAGGAACCTGTACGTATT

       1260        1270        1280        1290        1300
GCCACGGTCTGCTCTCTGTCCTACTCACAGCACATCCGCAGAGGCACACA
CGGTGCCAGACGAGAGACAGGATGAGTGTCGTGTAGGCGTCTCCGTGTGT

       1310        1320        1330        1340        1350
CGTACACCACTCTTCTATCGCTACTGCACGTCAGTGGCTGAGTCTGAGCA
GCATGTGGTGAGAAGATAGCGATGACGTGCAGTCACCGACTCAGACTCGT

      1360
ACTGTATGC
TGACATACG
```

homology with OY1

APPENDIX 3 (*contd.*): DNA SEQUENCE OF BRY.3B

```
        10         20         30         40         50
GCTACTGCACTTCCAGTGGGGCTGGAAGTGCTGGAGGCATGCCATGCTCA
GCATGACGTGAAGGTCACCCCGACCTTCACGACCTCCGTACGGTACGAGT
```

1st region of homology with OY1 begins (70% identity)

```
        60         70         80         90        100
CACCCTGTGTCTCCTAGGATGCCTCTGGTTTTTCATAGACAGCCTACCTT
GTGGGACACAGAGGATCCTACGGAGACCAAAAAGTATCTGTCGGATGGAA

       110        120        130        140        150
AGCTAATCTTGCGTTTTGTGTTGCATCGTGCCAATATGTGCCACCCTCAT
TCGATTAGAACGCAAAACACAACGTAGCACGGTTATACACGGTGGGAGTA

       160        170        180        190        200
GGAGACAGCATGCTGGCTGTTGGGCTGATCCTGGGCACGACAGGCCCAGG
CCTCTGTCGTACGACCGACAACCCGACTAGGACCCGTGCTGTCCGGGTCC

       210        220        230        240        250
ACCTTCCCAGTTGTCTACTGTGTGGAACAATCTCCAGTGCAGCAGGCAAT
TGGAAGGGTCAACAGATGACACACCTTGTTAGAGGTCACGTCGTCCGTTA

       260        270        280        290        300
CTGTGCTCTTTCTCGTTCAGGTCTCTGCCCTTTTCTGCAGGACCTTAGCC
GACACGAGAAAGAGCAAGTCCAGAGACGGGAAAAGACGTCCTGGAATCGG

       310        320        330        340        350
TCTCACACATCCATCTGCATAGCATATCACATTCACAGCTGCCACATGCG
AGAGTGTGTAGGTAGACGTATCGTATAGTGTAAGTGTCGACGGTGTACGC

       360        370        380        390        400
CCAACTGAGAGCTCACAGACAGTGTTCAAGCCTAACCACTGATTGTCTGA
GGTTGACTCTCGAGTGTCTGTCACAAGTTCGGATTGGTGACTAACAGACT

       410        420        430        440        450
GTGAGTCTCAGACGATGCACTCCTTGTCCAAACACGGTACAACAACACGG
CACTCAGAGTCTGCTACGTGAGGAACAGGTTTGTGCCATGTTGTTGTGCC

       460        470        480        490        500
TAGAAAATGGTGAGCATGTTTTTGCATTTGGTCTAATTGCAGATTTCTGG
ATCTTTTACCACTCGTACAAAAACGTAAACCAGATTAACGTCTAAAGACC
```

```
      510        520        530        540        550
AGCGACGGTCTGGAGTTACCACTGGCCATCCTTTCGTCGTTTCCATCCTG
TCGCTGCCAGACCTCAATGGTGACCGGTAGGAAAGCAGCAAAGGTAGGAC

      560        570        580        590        600
TACAGGGTCACTGCAGAGATACGGCGACCAACCACCCTTAGACTGGGGAC
ATGTCCCAGTGACGTCTCTATGCCGCTGGTTGGTGGGAATCTGACCCCTG

      610        620        630        640        650
GCCTCGTGCCTCCTTTTATGGTTGCTCAGTACCTCCTCTCCTGTCGCCTT
CGGAGCACGGAGGAAAATACCAACGAGTCATGGAGGAGAGGACAGCGGAA

      660        670        680        690        700
CATCTCTGTGTCTGCCCTGGGCCAGACACACACACACACACACACACACA
GTAGAGACACAGACGGGACCCGGTCTGTGTGTGTGTGTGTGTGTGTGTGT
                    (AC).(TG)15

      710        720        730        740        750
CACACATACACAAACACAGCTGACACAAGCACACACATGTATACACACAC
GTGTGTATGTGTTTGTGTCGACTGTGTTCGTGTGTGTACATATGTGTGTG

      760        770        780        790        800
ATCTCGCACTGGGACACAAATGCAAATAGGAAACAATGGTGTTTCAGAAG
TAGAGCGTGACCCTGTGTTTACGTTTATCCTTTGTTACCACAAAGTCTTC
                1st region of homology with OY1 ends

      810        820        830        840        850
CCCGTGTTCCCTGTTTGTGTTTTTGTGTCCGTGCCGAGCATGTATGTGTA
GGGCACAAGGGACAAACACAAAAACACAGGCACGGCTCGTACATACACAT

      860        870        880        890        900
TACACGCATGTCTTGTGTTGGCTGTGTGTGTGTGTGTGTGTGTATACACA
ATGTGCGTACAGAACACAACCGACACACACACACACACACACATATGTGT
                    (TG).(AC)10

      910        920        930        940        950
CGCATGTCTGTGTCGGCCCAGGAAGATGCAGAGATGAAGGTGACAGGAGA
GCGTACAGACACAGCCGGGTCCTTCTACGTCTCTACTTCCACTGTCCTCT

      960        970        980        990       1000
GAAGGTGCCAACCTACCATAAGAGAAGGCATGAGGCCTCCCAGTCTAAGG
CTTCCACGGTTGGATGGTATTCTCTTCCGTACTCCGGAGGGTCAGATTCC
```

1010 1020 1030 1040 1050
GTGGGTTGGGTTGCTGCAATGTTGGAGATGAAGTCGGTAGGCACCATGCC
CACCCAACCCAACGACGTTACAACCTCTACTTCAGCCATCCGTGGTACGG

2nd region of homology with OY1 begins (73% identity)

1060 1070 1080 1090 1100
CTGCCCTTGTGGGCTCCTGTCACTTCTCCCAAAGCTGATATAGGCAGCAG
GACGGGAACACCCGAGGACAGTGAAGAGGGTTTCGACTATATCCGTCGTC

1110 1120 1130 1140 1150
TCCAATATTTTCCCAAGGCAGTGTGAAGCCCCTTCCAGCGCAGATGTGTG
AGGTTATAAAAGGGTTCCGTCACACTTCGGGGAAGGTCGCGTCTACACAC

1160 1170 1180 1190 1200
TTGGTGGTGAGCCACTAAGTAGTGCATGTGGAAACTCACAAAAGATCCCA
AACCACCACTCGGTGATTCATCACGTACACCTTTGAGTGTTTTCTAGGGT

1210 1220 1230 1240 1250
AAGTCAGCACAAAGGTGCACTGGGACCCCAGAGAGGAAGCCCTCTCAATG
TTCAGTCGTGTTTCCACGTGACCCTGGGGTCTCTCCTTCGGGAGAGTTAC

1260 1270 1280 1290 1300
ACAGACCTTCCCAGGGTAAGGGCAGGAGTGAAGGCCCCAGAAAATCTCAG
TGTCTGGAAGGGTCCCATTCCCGTCCTCACTTCCGGGGTCTTTTAGAGTC

1310 1320 1330 1340 1350
TCGCAGTGAGCAACAGGAGGCCAGTTGATGGCAGGACCCATGGAAGGATG
AGCGTCACTCGTTGTCCTCCGGTCAACTACCGTCCTGGGTACCTTCCTAC

1360 1370 1380 1390 1400
ATTCTAACAAACATATTTAATTCACCGTGCCTCTAGGAATATCACCTAGG
TAAGATTGTTTGTATAAATTAAGTGGCACGGAGATCCTTATAGTGGATCC

1410 1420 1430 1440 1450
TCTCTTCACGCTGCAACCACAACTGATTTGAGGGAGACTTGCTTCACTCA
AGAGAAGTGCGACGTTGGTGTTGACTAAACTCCCTCTGAACGAAGTGAGT

1460 1470 1480 1490 1500
CAGACACACTGAATTCCCAAGCCCCAGTCTCAGCAGGCCACCTGCTACAT
GTCTGTGTGACTTAAGGGTTCGGGGTCAGAGTCGTCCGGTGGACGATGTA

```
      1510        1520        1530        1540        1550
GCATGGCCCAAATTCATACTCCGCAGCCAGCTTGCATGAGCCTGAGCATC
CGTACCGGGTTTAAGTATGAGGCGTCGGTCGAACGTACTCGGACTCGTAG

      1560        1570        1580        1590        1600
CACATTTGGGAAAGGACAGTGTCACGGCTCCAACGGCTAACCCTTCATGG
GTGTAAACCCTTTCCTGTCACAGTGCCGAGGTTGCCGATTGGGAAGTACC

      1610        1620        1630        1640        1650
CCCACCTCCAGACCCCTTCACCACATATGTCTCTCGTCAGTTGCCTCTCA
GGGTGGAGGTCTGGGGAAGTGGTGTATACAGAGAGCAGTCAACGGAGAGT

      1660        1670        1680        1690        1700
TGGAAGAATCTTCCCTCCTGGGGTAGTACTTCAGGGGATCGTCCACTCGT
ACCTTCTTAGAAGGGAGGACCCCATCATGAAGTCCCCTAGCAGGTGAGCA

      1710        1720        1730        1740        1750
CTTAGCTGATGATCTGCTGGCAACTACCGAGCCAATGTCAGTCCTGGGCT
GAATCGACTACTAGACGACCGTTGATGGCTCGGTTACAGTCAGGACCCGA

      1760        1770        1780        1790        1800
GACCAGGACCCAAACCCTCCCATGAGCAGCCAAGAACTCCAACATCAATC
CTGGTCCTGGGTTTGGGAGGGTACTCGTCGGTTCTTGAGGTTGTAGTTAG

      1810        1820        1830        1840        1850
CATCCAACTGTATCCAGGACCAAAAACAACCTCCACCTCAGCAATTCCGT
GTAGGTTGACATAGGTCCTGGTTTTTGTTGGAGGTGGAGTCGTTAAGGCA

      1860        1870        1880        1890        1900
GTTGATCCTGGGCAGTTGTGGCTGACAACCAGTTGAGAAAGTTTAGGCTC
CAACTAGGACCCGTCAACACCGACTGTTGGTCAACTCTTTCAAATCCTAG

      1910        1920        1930        1940        1950
CTGGTGTCAGCCTGTGGCTGCTCCTCCCTGTTCAAAGTCCCAGAACCAAT
GACCACAGTCGGACACCGACGAGGAGGGACAAGTTTCAGGGTCTTGGTTA

      1960        1970        1980        1990        2000
GGACTGGTGTGGAACGATGTGCCCTTTACCCTGCAGAAAGATAGGGGAGA
CCTGACCACACCTTGCTACACGGGAAATGGGACGTCTTTCTATCCCCTCT
```

2010
AGGGTGCGGATC
TCCCACGCCTAG
***Bam* HI**

APPENDIX 4: SEQUENCE OF cBRY.3

```
        10         20         30         40         50
CCCATATATATATATATAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGA
GGGTATATATATATATATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT
Sma I                  (AT)7(AG)17 repeat

        60         70         80         90         100
GTGTCAGAAATATAAACCCTCACCCTCTGGTCAGTTGATTCTTGGTAAAG
CACAGTCTTTATATTTGGGAGTGGGAGACCAGTCAACTAAGAACCATTTC

       110        120        130        140        150
GTTCAAGATCATTCAATTGGGGAAGAATAGTTTTTTCAACAAGTGTGCTG
CAAGTTCTAGTAAGTTAACCCCTTCTTATCAAAAAAGTTGTTCACACGAC

       160        170        180        190
GGACATTTGGATATTCACATGCAAAGAGTTAAATCAGGG
CCTGTAAACCTATAAGTGTACGTTTCTCAATTTAGTCCC
                                      Sma I
```

RESPONSE TO EXAMINERS COMMENTS:

CHAPTER 3, page 45: As the examiner notes, fragments 2B and 2C hybridize to the same 2.2 kb band in the *Hind III* digest. There is a problem with the markers in these photographs in that the prints of the Southern blots are slightly larger than the prints of the original gel, thus changing the position of the size markers slightly. BRY.2B and BRY.2C hybridize to a male-specific band to which BRY.2D shows no homology (the 2.2 kb *Hind III* band).

Section 3.34: The examiner is concerned about the copy number calculation used to estimate that approximately 40% of the bovine Y chromosome consists of sequences which are related to BRY.2 and BRY.3. This calculation has been made on the basis that there are about 1100 male-specific copies of BRY.2 (3.7 kb) and about 600 male-specific copies of BRY.3 (4.2 kb), which gives about 6.6×10^6 bp of a Y-chromosomal complement of 1.5×10^7 bp = 44%. BRY.2A is about 0.75 kb long and is a component of only some copies of BRY.2.

CHAPTER 5: The examiner asks whether nearly all of the hybridization of cDNA to the phage clones could be explained by the presence of transcribed interspersed repeated sequences. This in fact was the hypothesis I put forward in this Chapter (pages 73, 76 and 77) to explain the rather unexpected hybridization of labelled cDNA to many clones containing repeated sequences. The hypothesis was tested within this chapter by probing Southern blots of the restricted phage DNAs with nick-translated female DNA and in Chapter 6 by the isolation of a cDNA clone from a library using BRY.3. On pages 77 and 78 of Chapter 5 I drew attention to the fact that some phage clones which hybridized with the cDNA probes did not appear to contain interpersed genomic repeats (ie they did not hybridize with the female genomic probe).